MOLECULAR HOMOLOGY BETWEEN HUMAN AND DOG ON MAMMARY AND ORAL CANCERS

by

DELI LIU

(Under the Direction of Shaying Zhao)

ABSTRACT

Spontaneously occurring dog cancers represent excellent models of human cancers but are greatly understudied. To better utilize this valuable resource, we conducted the first comprehensive genome-wide characterization of dog mammary and oral cancers to evaluate dog-human molecular homology.

For the dog mammary cancer, we performed whole genome sequencing, whole exome sequencing, RNA-seq and/or high density arrays on 12 canine mammary cancer cases, including 7 simple carcinomas and 4 complex carcinomas. Canine simple carcinomas, which histologically match human breast carcinomas, harbor extensive genomic aberrations, many of which faithfully recapitulate key features of human breast cancer. Canine complex carcinomas, which are characterized by proliferation of both luminal and myoepithelial cells and are rare in human breast cancer, appear to lack genomic abnormalities. Instead, these tumors have about 35 chromatin-modification genes downregulated, and are abnormally enriched with active histone modification H4-acetylation while aberrantly depleted with repressive histone modification

H3K9me3. Our findings indicate the likelihood that canine simple carcinomas arise from genomic aberrations whereas complex carcinomas originate from epigenomic alterations, reinforcing their unique value. Canine complex carcinomas offer an ideal system to study myoepithelial cells, the second major cell lineage of the mammary gland. Canine simple carcinomas, which faithfully represent human breast carcinomas at the molecular level, provide indispensable models for basic and translational breast cancer research.

For the dog oral cancer, we investigated 12 canine head and neck squamous cell carcinoma (HNSCC) cases, of which 9 are oral, via high density arrays and RNA-seq. The analyses reveal that these canine cancers faithfully recapitulate many key molecular features of human HNSCC. These include similar genomic copy number abnormality landscapes, analogous sequence mutation patterns, and recurrent alteration of known HNSCC genes (e.g., *MYC*, *CDKN2A)* and pathways (e.g., cell cycle, mitogenic signaling, TGFβ signaling). Amplification or overexpression of protein kinase genes, matrix metalloproteinase genes, and epithelial–mesenchymal transition genes *TWIST1* and *SNAI* are also prominent in these canine tumors. Our study reemphasizes the value of spontaneous canine cancers in HNSCC basic research and anticancer drug discovery.

INDEX WORDS: Spontaneous canine mammary cancer; dog-human molecular homology; simple versus complex carcinomas; genomic versus epigenomic abnormality; luminal versus myoepithelial cells; spontaneous canine HNSCC; oral SCC; translational model; drug discovery.

# MOLECULAR HOMOLOGY BETWEEN HUMAN AND DOG ON MAMMARY AND ORAL CANCERS

by

DELI LIU

**B.S.**, China University of Petroleum (East China), China, 2007

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2014

MOLECULAR HOMOLOGY BETWEEN HUMAN AND DOG ON MAMMARY AND ORAL CANCERS

by

DELI LIU

| | |
|---|---|
| Major Professor: | Shaying Zhao |
| Committee: | Liming Cai |
| | Kevin Dobbin |
| | Travis Glenn |

Electronic Version Approved:

Julie Coffield
Interim Dean of the Graduate School
The University of Georgia
December 2014

DEDICATION

This dissertation is dedicated to my love for science.

ACKNOWLEDGEMENTS

I am grateful to my advisor, Professor Shaying Zhao for her support and guidance over the years. She patiently provided the vision, encouragement and advice necessary for me to proceed through the doctoral program and complete my dissertation. Without her guidance and persistent help this dissertation would not have been possible. I would also like to thank the members of my advisory committee: Drs. Liming Cai, Kevin Dobbin and Travis Glenn for their mentorship and guidance during my graduate training. I would like to thank all past and present members of the Zhao lab for their friendship and help. I would like to thank my wife Huimin for standing beside me throughout my Ph.D. studies.

TABLE OF CONTENTS

Page

CHAPTER 1

# INTRODUCTION AND LITERATURE REVIEW

## INTRODUCTION TO CANCER

### What is cancer?

Cancer is a group of disease with abnormally growing cells, which will potentially invade to other parts of body (Horsfall 1963). Although there are many kinds of cancer, all cancers start because abnormal cells grow out of control. Cancer develops when the normal regulatory mechanisms, differentiation and death mechanisms are disrupted. Untreated cancers can cause serious illness and death.

### Cancer prevalence

Cancer is a major public health problem in the United States and many other parts of the world. One in 4 deaths in the United States is due to cancer (Siegel et al. 2014). A total of 1,665,540 new cancer cases and 585,720 cancer deaths are projected to occur in the United States in 2014 (Siegel et al. 2014), and up to 2.3 million cancer cases in 2030 (Smith et al. 2009).

## CANCER GENOME SEQUENCING

Cancer genome sequencing is the sequencing of a single, homogeneous or heterogeneous group of cancer cells. It is a biochemical laboratory method for the characterization and

identification of the DNA or RNA sequences of cancer cells. After the human genome project finished in 2003 (Consortium. 2004), next generation sequencing has become the essential tool in cancer studies, diagnostics as and targeted therapy. Cancer genome sequencing is not limited to whole-genome sequencing (WGS), and can also include whole-exome sequencing (WES) and transcriptome sequencing (RNA-seq). These methods can be used to detect not only the genomic changes, e.g., point mutations, small indels, copy number abnormalities, translocations, but also the transcriptomic changes, e.g., gene expression alterations, fusion genes and alternative splicing events.

The Cancer Genome Atlas (TCGA) is a project, begun in 2005, to catalogue genetic mutations responsible for cancer, using genome sequencing and bioinformatics (http://cancergenome.nih.gov/), with an investment of $50 million each from the National Cancer Institute (NCI) and National Human Genome Research Institute (NHGRI). By October 2014, TCGA has included 34 different tumor types, with hundreds of samples in each tumor type (Network. 2011; Network. 2012a; Network. 2012b; Network. 2013; de Castro and Negrao 2014; Network. 2014). The project uses many different techniques to analyze the patient samples, e.g., WGS, WES, RNA-seq, SNP array, gene expression array and DNA methylation.

## MUTATIONS IN CANCER

### Cancer is a disease of the genome

Cancer is a disease of the genome, and thousands of somatic sequence mutations and hundreds of chromosomal aberrations could be found in a typical cancer genome (Macconaill

and Garraway 2010; Alexandrov et al. 2013). As of October 2014, a total of 522 genes have been found as known cancer driver mutations in Cancer Gene Census (CGC) database (http://cancer.sanger.ac.uk/cancergenome/projects/census/). Based on the CGC data, most of somatic mutations (90%) are oncogenetic and only 10% are tumor suppressors.

**Sequencing mutations in cancer**

In the cancer genome, somatic mutations usually include base substitutions, insertions and deletions or structural variations (Helleday et al. 2014). The first report of cancer genome sequencing appeared in 2006 (Ley et al. 2008). In that study, 13,023 genes were sequenced in 11 breast and 11 colorectal tumors. After that, the same group added over 5,000 more genes and almost 8,000 transcripts to complete the exomes of 11 breast and colorectal tumors in 2007 (Shah et al. 2009). Due to the low cost of next-generation sequencing (NGS) technique in recent years, many types of cancers with hundreds of samples have been sequenced so far, and many new cancer genes have been identified (Banerji et al. 2012; Barbieri et al. 2012; Network. 2012a; Network. 2012b; Stephens et al. 2012; Alexandrov et al. 2013; Network. 2014).

**HUMAN BREAST CANCER**

Breast cancer is the development of cancer from breast tissue in humans. The most common symptom of breast cancer is a new lump or mass. Other symptoms of breast cancer may include swelling of the breast, skin dimpling, or redness of breast skin (Nguyen et al. 2011). Breast cancer occurs in both men and women, although male breast cancer is rare. In 2014, there will be 232,670 new cases of breast cancer in females, 62,570 new cases of carcinoma in situ

(non-invasive) and 40,000 will die from breast cancers the United States (Siegel et al. 2014). Most studies divide breast cancer into five major molecular subtypes (Prat and Perou 2011): luminal A (40%), luminal B (20%), basal-like (15-20%), HER2-enriched (10-15%) and normal-like (6-10%). Breast cancer is one of most well-studied cancer types in human, especially in females. By October 2014, more than 1,000 samples of human breast invasive carcinomas have been analyzed in TCGA project, with WGS, WES, SNP array, DNA methylation, mRNA expression array, RNA-seq and microRNA techniques (Network. 2012b). Many genes (e.g., TP53, PTEN, AKT1, PIK3CA and GATA3) have been found altered in human breast cancer (Banerji et al. 2012; Network. 2012b; Stephens et al. 2012).

**HUMAN ORAL CANCER**

Head and neck cancers are the cancers that usually start in the oral cavity, nasal cavity, paranasal sinuses, pharynx, and larynx, and 90% of head and neck cancers are squamous cell carcinomas (HNSCC) (Sanderson and Ironside 2002). HNSCC is one of the most common cancers worldwide, with annual incidence rate of 30/100, 000 in India and France (Sanderson and Ironside 2002). Oral squamous cell carcinoma (oSCC), a subset of HNSCC, has a 5-year survival rate of only about 50% (Leemans et al. 2011; Pickering et al. 2013). In 2014, there will be 37,000 new cases of oral cancers, and 7,300 people will die of these cancers in the United States. Human oral cancer is one of the better studied cancers. By October 2014, more than 500 samples of human HNSCC have been analyzed in TCGA project, and many genes (e.g., TP53, CDKN2A, PIK3CA, HRAS and NOTCH1) have been found altered in human HNSCC and oSCC (Agrawal et al. 2011; Stransky et al. 2011).

## DOG CANCER

Cancer is the leading cause of death in dogs over the age of 10. Almost 50% of dogs over the age of 10 will develop cancers, and about one in four of all dogs will develop a cancer in their lifetime (Silva et al. 2013). In the basic research, dog cancer has become the great model to study many types of human cancers (Ji et al. 2010; Beck et al. 2013; Tang et al. 2014). In the clinical research, dog cancers could provide the valuable resource to identify the cancer-associated genes, understand the tumor biology and develop novel cancer therapeutics (Paoloni and Khanna 2008).

## DOG MAMMARY CANCER

As in women, mammary cancer is the most frequently diagnosed cancer in female dogs, accounting for 70% of all cancer cases (Merlo et al. 2008). Table 1 shows that the annual incidence rate of dog mammary cancer is estimated at 192/100,000 (Merlo et al. 2008), which is comparable to the rate of 125/100,000 for breast cancer in women in the United States (Siegel et al. 2014). Dog mammary cancers are more common in intact than in spayed females. The risk of developing mammary cancer is significantly lowered by spaying at an early age, resulting in lower incidence rates in countries where this surgery is the common practice (Beck et al. 2013).

## DOG ORAL CANCER

The oral cavity is a common site for malignant tumors, accounting for 5% to 7% of all canine cancers, and about 17% to 25% are squamous cell carcinomas (Oakes et al. 1993). Table

1 shows that the annual incidence rate of dog oral cancers is about 20/100,000 (Fulton et al. 2013), which is comparable to the rate of 12/100,000 for human oral cancer in the United States (Siegel et al. 2014). Dog oral cancers grow very quickly, involving the bone and tissue near them, and metastasize quickly and easily to other areas of the body. Oral squamous cell carcinoma occurs most commonly in the gingiva, particularly rostral to the canine teeth (Todoroff and Brodey 1979).

## EXPERIMENTAL DESIGN AND INNOVATION

### Spontaneously occurring canine mammary cancer represents an excellent model of human breast cancer

Unlike traditional mouse models, dog cancer serves as a great model analogous to humans for several reasons:

- First, dog cancers are naturally occurring and heterogeneous, capturing the essence of human cancer, unlike genetically modified or xenograft rodent models. As companion animals, dogs share the human environment and are exposed to many of the same carcinogens. Indeed, environmental toxins, advancing age, and obesity are also risk factors for canine cancer (Meuten 2002). For example, canine tonsillar tumors are more common in large cities but rare in rural areas (Reif and Cohen 1971).

- Second, dogs better resemble humans in biology, e.g., similar telomere and telomerase activities (Nasir et al. 2001) and frequent spontaneous epithelial cancers (Meuten 2002), unlike mice (Rangarajan and Weinberg 2003). Critically, numerous anatomic and

clinical similarities are noted for the same types/subtypes of cancer between the two species and similar treatment schemes have been used (Paoloni and Khanna 2008; Gordon et al. 2009; Rowell et al. 2011).

- Third, the large population of pet dogs (~70 million estimated in the US) provides a valuable resource facilitating basic and clinical research. Dogs receive a high level of healthcare with over one billion dollars per year spent on dog healthcare in the United States (American Veterinary Medical Association, 2013). In 2013, about 45% of dogs are more than six years old, which is equivalent to 60–95 years of age in humans. Thus, the dog could be a great model to study genetic and environmental factors in human cancer, which primarily is a disease of aging (Howlader et al. 2010).

**Different from human cancers, dog cancers are poorly characterized at the genome-wide level.**

Unlike human breast and oral cancer, dog mammary cancer (MC) and oral cancer remain largely uncharacterized at the molecular level. For the dog mammary cancer, only five dog MC cases have recently undergone ~2X whole genome sequencing (Beck et al. 2013), and a limited number have been analyzed with gene expression microarray (Rao et al. 2009; Klopfleisch et al. 2010; Paw Owski et al. 2013). For the dog oral cancer, not a single canine HNSCC genome or transcriptome has ever been investigated by sequencing, microarray, or other strategies. This drastically differs from their human counterparts, where thousands of breast cancer genomes and transcriptomes are characterized (Network. 2012b), and hundreds of human oSCC genomes and transcriptomes have been studied.

**Our study to investigate the molecular homology between human and dog on mammary and oral cancers**

Because there are insufficient data to evaluate the molecular similarities between these dog cancers and their human counterparts, a key factor to consider in assessing the usage of spontaneous canine cancers in anticancer drug development, we set out to conduct the first comprehensive genome-wide characterization of dog mammary and oral cancers to evaluate the dog-human molecular homology.

To test the molecular similarities between human and dog on mammary and oral cancers, first, we performed whole genome sequencing, whole exome sequencing, RNA-seq and/or high density arrays on 12 dog mammary cancers, and found that dog simple carcinoma, one subtype of dog mammary cancer, represents human breast cancer at the molecular level. Second, we investigated 12 dog HNSCC cases, of which 9 are oSCC, via high density arrays and RNA-seq analyses, and revealed these dog oral cancers faithfully recapitulate many key molecular features of human oSCC.

**TABLE LEGEND**

**Table 1. The annual incidence rate of human and dog cancers.**

| | Annual Incidence per 100,000 | |
|---|---|---|
| Cancer Type | Human | Dog |
| Breast | 125 (Siegel et al. 2014) | 192 (Merlo et al. 2008) |
| Oral | 12 (Siegel et al. 2014) | 20 (Fulton et al. 2013) |
| Skin | 40 (Siegel et al. 2014) | 19 (Merlo et al. 2008) |
| Lung | 90 (Siegel et al. 2014) | 4 (Conti et al. 2010) |
| Lymphoma | 40 (Siegel et al. 2014) | 20 (Wiedemann et al. 2005) |
| Bone | 0.95 (Siegel et al. 2014) | 4 (Anfinsen et al. 2011) |

## REFERENCES


Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J et al. 2011. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* **333**(6046): 1154-1157.

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL et al. 2013. Signatures of mutational processes in human cancer. *Nature* **500**(7463): 415-421.

Anfinsen KP, Grotmol T, Bruland OS, Jonasdottir TJ. 2011. Breed-specific incidence rates of canine primary bone tumors--a population based survey of dogs in Norway. *Canadian journal of veterinary research = Revue canadienne de recherche veterinaire* **75**(3): 209-215.

Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, Lawrence MS, Sivachenko AY, Sougnez C, Zou L et al. 2012. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* **486**(7403): 405-409.

Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M, Theurillat JP, White TA, Stojanov P, Van Allen E, Stransky N et al. 2012. Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer. *Nature genetics* **44**(6): 685-689.

Beck J, Hennecke S, Bornemann-Kolatzki K, Urnovitz HB, Neumann S, Strobel P, Kaup FJ, Brenig B, Schutz E. 2013. Genome aberrations in canine mammary carcinomas and their detection in cell-free plasma DNA. *PloS one* **8**(9): e75485.

Consortium. IHGS. 2004. Finishing the euchromatic sequence of the human genome. *Nature* **431**(7011): 931-945.

Conti MB, Marchesi MC, Angeli G, Lepri E, Marinetti C, Rueca F. 2010. A case of primary papillary disseminated adenocarcinoma of canine lung. *Veterinary research communications* **34 Suppl 1**: S111-115.

de Castro G, Jr., Negrao MV. 2014. The Cancer Genome Atlas findings in head and neck cancer: a renewed hope. *Current opinion in oncology* **26**(3): 245-246.

Fulton AJ, Nemec A, Murphy BG, Kass PH, Verstraete FJ. 2013. Risk factors associated with survival in dogs with nontonsillar oral squamous cell carcinoma 31 cases (1990-2010). *Journal of the American Veterinary Medical Association* **243**(5): 696-702.

Gordon I, Paoloni M, Mazcko C, Khanna C. 2009. The Comparative Oncology Trials Consortium: using spontaneously occurring cancers in dogs to inform the cancer drug development pathway. *PLoS medicine* **6**(10): e1000161.

Helleday T, Eshtad S, Nik-Zainal S. 2014. Mechanisms underlying mutational signatures in human cancers. *Nature reviews Genetics* **15**(9): 585-598.

Horsfall FL, Jr. 1963. Current Concepts of Cancer. *Canadian Medical Association journal* **89**: 1224-1229.

Howlader N, Ries LA, Mariotto AB, Reichman ME, Ruhl J, Cronin KA. 2010. Improved estimates of cancer-specific survival rates from population-based data. *Journal of the National Cancer Institute* **102**(20): 1584-1598.

Ji X, Tang J, Halberg R, Busam D, Ferriera S, Pena MM, Venkataramu C, Yeatman TJ, Zhao S. 2010. Distinguishing between cancer driver and passenger gene alteration candidates via cross-species comparison: a pilot study. *BMC cancer* **10**: 426.

Klopfleisch R, Lenze D, Hummel M, Gruber AD. 2010. Metastatic canine mammary carcinomas can be identified by a gene expression profile that partly overlaps with human breast cancer profiles. *BMC cancer* **10**: 618.

Leemans CR, Braakhuis BJ, Brakenhoff RH. 2011. The molecular biology of head and neck cancer. *Nature reviews Cancer* **11**(1): 9-22.

Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M et al. 2008. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* **456**(7218): 66-72.

Macconaill LE, Garraway LA. 2010. Clinical implications of the cancer genome. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**(35): 5219-5228.

Merlo DF, Rossi L, Pellegrino C, Ceppi M, Cardellino U, Capurro C, Ratto A, Sambucco PL, Sestito V, Tanara G et al. 2008. Cancer incidence in pet dogs: findings of the Animal Tumor Registry of Genoa, Italy. *Journal of veterinary internal medicine / American College of Veterinary Internal Medicine* **22**(4): 976-984.

Meuten DJ. 2002. *Tumors in domestic animals*. Iowa State University Press, Ames, Iowa.

Nasir L, Devlin P, McKevitt T, Rutteman G, Argyle DJ. 2001. Telomere lengths and telomerase activity in dog tissues: a potential model system to study human telomere and telomerase biology. *Neoplasia* **3**(4): 351-359.

Network. CGA. 2012a. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**(7407): 330-337.

-. 2012b. Comprehensive molecular portraits of human breast tumours. *Nature* **490**(7418): 61-70.

-. 2013. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**(7456): 43-49.

-. 2014. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**(7511): 543-550.

Network. CGAR. 2011. Integrated genomic analyses of ovarian carcinoma. *Nature* **474**(7353): 609-615.

Nguyen J, Cramarossa G, Bruner D, Chen E, Khan L, Leung A, Lutz S, Chow E. 2011. A literature review of symptom clusters in patients with breast cancer. *Expert review of pharmacoeconomics & outcomes research* **11**(5): 533-539.

Oakes MG, Lewis DD, Hedlund CS, Hosgood G. 1993. Canine Oral Neoplasia. *Comp Cont Educ Pract* **15**(1): 15-31.

Paoloni M, Khanna C. 2008. Translation of new cancer treatments from pet dogs to humans. *Nature reviews Cancer* **8**(2): 147-156.

Paw Owski KM, Maciejewski H, Dolka I, Mol JA, Motyl T, Krol M. 2013. Gene expression profiles in canine mammary carcinomas of various grades of malignancy. *BMC veterinary research* **9**(1): 78.

Pickering CR, Zhang J, Yoo SY, Bengtsson L, Moorthy S, Neskey DM, Zhao M, Ortega Alves MV, Chang K, Drummond J et al. 2013. Integrative genomic characterization of oral squamous cell carcinoma identifies frequent somatic drivers. *Cancer discovery* **3**(7): 770-781.

Prat A, Perou CM. 2011. Deconstructing the molecular portraits of breast cancer. *Molecular oncology* **5**(1): 5-23.

Rangarajan A, Weinberg RA. 2003. Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nature reviews Cancer* **3**(12): 952-959.

Rao NA, van Wolferen ME, Gracanin A, Bhatti SF, Krol M, Holstege FC, Mol JA. 2009. Gene expression profiles of progestin-induced canine mammary hyperplasia and spontaneous mammary tumors. *Journal of physiology and pharmacology : an official journal of the Polish Physiological Society* **60 Suppl 1**: 73-84.

Reif JS, Cohen D. 1971. The environmental distribution of canine respiratory tract neoplasms. *Archives of environmental health* **22**(1): 136-140.

Rowell JL, McCarthy DO, Alvarez CE. 2011. Dog models of naturally occurring cancer. *Trends in molecular medicine* **17**(7): 380-388.

Sanderson RJ, Ironside JA. 2002. Squamous cell carcinomas of the head and neck. *BMJ* **325**(7368): 822-827.

Shah SP, Morin RD, Khattra J, Prentice L, Pugh T, Burleigh A, Delaney A, Gelmon K, Guliany R, Senz J et al. 2009. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature* **461**(7265): 809-813.

Siegel R, Ma J, Zou Z, Jemal A. 2014. Cancer statistics, 2014. *CA: a cancer journal for clinicians* **64**(1): 9-29.

Silva G, Fachin AL, Beleboni RO, Franca SC, Marins M. 2013. In vitro action of flavonoids in the canine malignant histiocytic cell line DH82. *Molecules* **18**(12): 15448-15463.

Smith BD, Smith GL, Hurria A, Hortobagyi GN, Buchholz TA. 2009. Future of cancer incidence in the United States: burdens upon an aging, changing nation. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **27**(17): 2758-2765.

Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR et al. 2012. The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**(7403): 400-404.

Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence MS, Sougnez C, McKenna A et al. 2011. The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**(6046): 1157-1160.

Tang J, Li Y, Lyon K, Camps J, Dalton S, Ried T, Zhao S. 2014. Cancer driver-passenger distinction via sporadic human and dog cancer comparison: a proof-of-principle study with colorectal cancer. *Oncogene* **33**(7): 814-822.

Todoroff RJ, Brodey RS. 1979. Oral and pharyngeal neoplasia in the dog: a retrospective survey of 361 cases. *Journal of the American Veterinary Medical Association* **175**(6): 567-571.

Wiedemann AL, Charney SC, Barger AM, Schaeffer DJ, Kitchell BE. 2005. Assessment of corticosteroid-induced alkaline phosphatase as a prognostic indicator in canine lymphoma. *The Journal of small animal practice* **46**(4): 185-190.

# CHAPTER 2

# MOLECULAR HOMOLOGY AND DIFFERENCE BETWEEN SPONTANEOUS CANINE MAMMARY CANCER AND HUMAN BREAST CANCER[1]

[1]Liu D, Xiong H, Ellis AE, Northrup NC, Rodriguez CO, Jr., et al. Molecular homology and difference between spontaneous canine mammary cancer and human breast cancer. 2014, Cancer Research 74(18):5045-5056.

**ABSTRACT**

Spontaneously occurring canine mammary cancer (MC) represents an excellent model of human breast cancer but is greatly understudied. To better utilize this valuable resource, we performed whole genome sequencing, whole exome sequencing, RNA-seq and/or high density arrays on 12 canine MC cases, including 7 simple carcinomas and four complex carcinomas. Canine simple carcinomas, which histologically match human breast carcinomas, harbor extensive genomic aberrations, many of which faithfully recapitulate key features of human breast cancer. Canine complex carcinomas, which are characterized by proliferation of both luminal and myoepithelial cells and are rare in human breast cancer, appear to lack genomic abnormalities. Instead, these tumors have about 35 chromatin-modification genes downregulated, and are abnormally enriched with active histone modification H4-acetylation while aberrantly depleted with repressive histone modification H3K9me3. Our findings indicate the likelihood that canine simple carcinomas arise from genomic aberrations whereas complex carcinomas originate from epigenomic alterations, reinforcing their unique value. Canine complex carcinomas offer an ideal system to study myoepithelial cells, the second major cell lineage of the mammary gland. Canine simple carcinomas, which faithfully represent human breast carcinomas at the molecular level, provide indispensable models for basic and translational breast cancer research.

## INTRODUCTION

Spontaneous cancers in pet dogs represent one of the best cancer models (Meuten 2002; Paoloni and Khanna 2008; Gordon et al. 2009; Parker et al. 2010; Tang et al. 2010; Rowell et al. 2011; Youmans et al. 2012; Tang et al. 2013). First, these cancers are naturally-occurring and heterogeneous, capturing the essence of human cancer, unlike genetically-modified or xenograft rodent models. Second, as companion animals, dogs share the same environment as humans and are exposed to many of the same carcinogens. Indeed, environmental toxins, advancing age and obesity are also risk factors for canine cancer (Meuten 2002). Third, dogs better resemble humans in biology, e.g., similar telomere and telomerase activities (Nasir et al. 2001) and frequent spontaneous epithelial cancers (Meuten 2002), unlike mice (Rangarajan and Weinberg 2003). Fourth, numerous anatomic and clinical similarities are noted for the same types/subtypes of cancer between the two species, and similar treatment schemes are used (Paoloni and Khanna 2008; Gordon et al. 2009; Rowell et al. 2011). Furthermore, the large population of pet dogs (~70 million estimated in the US) provides a valuable resource facilitating basic and clinical research. Importantly, the dog genome has been sequenced to a >7.6X coverage (Lindblad-Toh et al. 2005), yielding a genome assembly nearly as accurate as the mouse or rat genome (Lindblad-Toh et al. 2005; Ji and Zhao 2008), unlike another companion animal, the cat. This makes many genomic analyses possible with the dog but not with the cat.

As in women, mammary cancer (MC) is among the most frequent cancers in female dogs. The annual incidence rate is estimated at 198/100,000 (Meuten 2002),

which is comparable to the rate of 125/100,000 for breast cancer in women in the US (Siegel et al. 2012). MC is especially common in dogs that are not spayed or are spayed after the second estrus, with the risk for malignant tumor development expected at 26% (Meuten 2002). However, unlike human breast cancer, canine MC is poorly characterized at the genome-wide level. For example, only five canine MC cases have recently undergone ~2X whole genome sequencing (WGS) (Beck et al. 2013), and a limited number have been analyzed with gene expression microarray (Rao et al. 2009; Klopfleisch et al. 2010; Paw Owski et al. 2013). This drastically differs from their human counterparts, where thousands of breast cancer genomes and transcriptomes are characterized, with several studies cited here (Naylor et al. 2005; Sjoblom et al. 2006; Banerji et al. 2012; Cancer Genome Atlas 2012; Curtis et al. 2012; Stephens et al. 2012).

Like other cancer types, many anatomic and clinical similarities are documented between canine MC and human breast cancer (Sleeckx et al. 2011). Various molecular homologies are also reported, e.g., WNT signaling alteration (Rao et al. 2009; Klopfleisch et al. 2010). Meanwhile, canine MC also differs from human breast cancer in certain aspects. For example, dogs have only one or two estrous cycles a year followed by a prolonged luteal phase, which is 63 days for the dog compared to 14 days for the human. During this unusually long luteal phase, the mammary gland is continuously exposed to high levels of progesterone (Sleeckx et al. 2011). Another variation that may be related to this hormonal difference is described below.

Mammary gland epithelium consists of an inner layer of luminal cells and an outer layer of myoepithelial cells that border the basal lamina. Compared to luminal cells, myoepithelial cells are significantly understudied and poorly characterized (Lakhani and O'Hare 2001; Adriance et al. 2005; Gudjonsson et al. 2005; Polyak and Hu 2005; Sopel 2010; Moumen et al. 2011). Although their importance in mammary gland development and pathogenesis has been noted (Gudjonsson et al. 2005; Polyak and Hu 2005; Hu et al. 2008), myoepithelial cells have traditionally received far less attention than luminal cells. This is at least partially because they rarely proliferate in human breast cancer (Hayes 2011; Tan and Ellis 2013). However, in canine MC, myoepithelial cell proliferation is much more common, occurring in >20% canine tumors compared to <0.1% human tumors (Goldschmidt et al. 2011). To more effectively utilize this unique feature of canine MC for a better understanding of myoepithelial cells, we set out to comprehensively compare spontaneous canine MCs with and without myoepithelial cell proliferation and to evaluate their molecular similarities to and differences from human breast cancers.

## RESULTS

### Canine simple carcinomas have no myoepithelial cell proliferation, whereas canine complex carcinomas have luminal and myoepithelial cells both proliferating.

The 12 cases subjected to genome-wide characterization represent two major histologic subtypes of canine MC (Goldschmidt et al. 2011), five with myoepithelial cell proliferation (complex carcinomas) and 7 without (simple carcinomas). Tumor cells in

simple carcinomas express only luminal markers such as E-cadherin (Figure 1A), and histologically match typical human breast carcinomas (Figures 1A and 1C). Tumors with myoepithelial cell proliferation include four complex carcinomas, a subtype that is rare in humans (32), and one carcinoma with two distinct histologic regions, one considered simple and the other considered complex. Complex carcinomas have prominent expression of both the luminal marker E-cadherin and the myoepithelial marker smooth muscle myosin heavy chain (SMHC) (Figure 1B), indicating dual proliferation of luminal and myoepithelial cells. This is also visible in H&E stained sections (Figure 1D). Besides this histologic difference, the tumors also vary in cancer progression stages (*in situ*, invasive or metastatic to the lung) and in estrogen receptor (ER) expression (five ER+ tumors and 7 ER- tumors).

**Copy number abnormalities (CNAs) are frequent in canine simple carcinomas.**

Reminiscent of human breast cancer, canine simple carcinoma genomes harbor extensive CNAs. First, we observed both focal and broad CNAs totaling from 10Mb to >100Mb and affecting hundreds of genes per tumor (Figures 2A and 2B), with the extensiveness of CNA in correlation with the tumor progression stage. The only exception to this is an inflammatory carcinoma, where no CNAs were detected. Second, CNAs also occurred in genomic sites recurrently altered in human breast cancer, e.g., amplification of human 8q and dog chromosome 13 that encode genes including *MYC* (Figure 2A). Third, while large deletions were discovered, one resulting in *PTEN* loss (Figure 2A), amplifications prevailed over deletions in most tumors. Notably, two large amplicons of >4Mb, harboring 54 and 43 genes respectively, were uncovered (Figure 2C).

This led to amplification and overexpression of oncogenes such as *BRAF*, *PIM1* and *CCND3*.

**Translocations and a superamplicon were discovered in a canine simple carcinoma by paired-end WGS.**

To further explore the two >4Mb amplicons described above, we sequenced the tumor and normal genomes of case 76 (Figure 2A) to a >15X sequence coverage. For comparison purposes, similar sequencing was performed on the case having the most extensive CNAs (case 406434, with pulmonary metastasis) and another case having hardly any CNAs (case 32510). WGS revealed fewer translocations and inversions than CNAs in these tumors. Furthermore, reminiscent of the human breast cancer MCF7 genome (Volik et al. 2003), some translocations are associated with amplification, creating a superamplicon with loci from different chromosomes co-localized and co-amplified (Figure 2D).

Based on chimeric sequence reads that span the translocation junctions and PCR confirmation, we propose a mechanism for the superamplicon formation (Figure 2D). First, a circle, consisting of ~1Mb sequences from chromosome 12 and ~0.4Mb from chromosome 16, emerged via two translocations that were likely facilitated by prior sequence amplification. The circle, which harbors oncogene *PIM1* and 17 other genes, was then further amplified.

**The superamplicon harbors a potentially oncogenic fusion gene, *ZFAND3-MGAM*, created via a translocation.**

The superamplicon also harbors a newly created fusion gene. It consists of the first four exons of *ZFAND3*, a zinc finger gene located on chromosome 12, and the last 22-49 exons of *MGAM*, which encodes maltase-glucoamylase and is located on chromosome 16 (Figure 2E). The fusion gene, termed *ZFAND3-MGAM*, arose from a translocation occurring in introns; transcription and splicing then yielded an in-frame fusion transcript. This was confirmed by the detection of chimeric sequence fusion points via WGS, RNA-seq and PCR analyses.

As a result of in-frame fusion, the A20-type zinc finger domain of ZFAND3 and the glucoamylase domain of MGAM are kept intact in the fusion product (Figure 2F). This appears to be significant. First, the A20 zinc finger protein has been reported to inhibit tumor necrosis factor-induced apoptosis (Lademann et al. 2001). Second, MGAM, an integral membrane protein with its catalytic domains facing the extracellular environment, is normally expressed in the intestine to digest starch into glucose (Sauer et al. 2000). Indeed, we did not detect *MGAM* expression in normal mammary tissues, unlike *ZFAND3*. However, in carcinoma 76, both *MGAM* and *ZFAND3-MGAM* are amplified and overexpressed. ZFAND3-MGAM, which lacks the transmembrane domain and becomes intracellular, could promote oncogenesis by accelerating carbohydrate-metabolism via its glucoamylase domain and meanwhile inhibiting apoptosis via its A20-type domain. Of course, whether this is true or not requires further studies.

**Somatic sequence mutations are frequent in canine simple carcinomas as revealed by WES.**

To examine somatic base substitutions and small indels, we performed WES on the matching tumor and normal genomes of four simple carcinoma cases to 134-245X coverage. This analysis again revealed several dog-human homologies. First, base transitions, particularly C→T/G→A changes, dominate base transversions in most tumors (Figure 3A), indicating similar mutation mechanisms (e.g., deamination of 5mC to T) in both species. The only exception (tumor 406434) has C→A/G→T transversions predominating, which is not an experimental artifact of WES (Costello et al. 2013) based on our analyses, and concurrently harbors an altered *POLD1*. This likewise is consistent with human cancer studies that link C→A/G→T changes to *POLD1* mutations (Palles et al. 2012). Second, the mutation rate varies greatly among the carcinomas, with tumor 5 having 907 genes significantly mutated, compared to 0-31 genes for tumors of similar or more advanced stages (Figure 3A). This hypermutation is likely linked to defective DNA repair as well, because tumor 5 has as many as 24 DNA repair-associated genes mutated. Third, many known human breast cancer genes are also mutated in these canine tumors, including *BRCA1*, *IGF2R*, *FOXC2*, *DLG2* and *USH2A* as described below.

*USH2A* is one of the most significantly mutated genes in our study (p = 2.78E-12), having one nonsense-, 12 missense- and three synonymous-mutations (Figure 3B). Critically, *USH2A* is also prominently mutated in human breast cancer, ranking as the 21$^{st}$ most significantly mutated gene in the Cancer Genome Atlas (TCGA) study (Cancer

Genome Atlas 2012). *USH2A* alterations may contribute to MC pathogenesis in both dogs and humans.

**Canine complex carcinomas have hardly any genomic CNAs and their sequence mutation rates also appear low.**

Unlike simple carcinomas, we observed very few genomic CNAs in complex carcinomas, making their genomes appear normal (Figure 2A). Their sequence mutation rates are also low, based on our analysis with RNA-seq data. Briefly, to achieve a more accurate mutation-finding, we utilized only regions with 30-300X RNA-seq read coverage, which distribute across the genome and amount to 4.5-9.4 Mb sequence per sample. The analysis indicates that the mutation rates of complex carcinomas are significantly lower than those of simple carcinomas, but are comparable to those of normal mammary gland tissues (Figure 3C).

**Numerous chromatin-modification genes are downregulated in canine complex carcinomas.**

RNA-seq analysis revealed 751 total genes differentially expressed at FDR $\leq 0.2$ between simple and complex carcinomas (Figure 3D). Strikingly, among these genes, chromatin modification and transcription regulation are the most significantly enriched functions. Indeed, a total of 35 known chromatin-modification genes were found to be downregulated in complex carcinomas (Figure 3E), and over 40% of them remain so at $p \leq 0.05$ when further compared to normal mammary gland tissues. Moreover, chromatin modification stays as the most significantly enriched function amid genes (327 in total)

differentially expressed among the three types of samples. The same overall conclusions were reached at FDR $\leq 0.1$.

Amid the 35 chromatin genes downregulated in complex carcinomas, 30 encode histone modifiers, covering methylation and demethylation; acetylation and deacetylation; and ubiquitination and deubiquitination (Figure 3E). Intriguingly, both active and repressive modifiers were noted (see the paragraph follows). Furthermore, the identified histone acetyltransferases and deacetylase modify histones H3, H4 and H2A, influencing not only gene transcription (e.g., CREBBP) but also chromatin packing (e.g., MSL1 on H4K16 acetylation) (Fraga et al. 2005). Besides histone-modification genes, other types of chromatin-remodeling genes were also found downregulated in complex carcinomas (Figure 3E), most of which (e.g., *ARID1B, ASF1A* and *DNMT3B*) are known to be mutated in human cancers (Wilson and Roberts 2011; Dawson and Kouzarides 2012).

To understand the significance of the observed change in chromatin-modification genes, many encoding histone modifiers (Figure 3E), we investigated histone modification. Specifically, we performed IHC experiments to examine H3K9me3, a repressive modification that is associated with gene silencing and heterochromatin and for which six relevant genes are downregulated in complex carcinomas. These include H3K9 methyltransferase genes *SETDB1*, *EHMT1*, *EHMT2* and *SUZ12*, along with demethylase genes *JMJD1C* and *PHF2*. Meanwhile, we also examined H4-acetylation because at least 8 of the downregulated genes (*CREBBP*, *CSRP2BP*, *ING3*, *KAT2A*, *MSL1*, *MSL2*, *NCOA3* and *SIRT1*) are involved in histone acetylation or deacetylation.

Another active modification, H3K4me3, was studied as well because H3K4 methyltransferase genes *SETD1A*, *MLL2* and *MLL4* are among those downregulated.

**In canine normal mammary glands, both active and repressive histone modifications are significantly depleted in myoepithelial cells when compared to luminal cells.**

To understand the alteration in cancer, we first investigated canine normal mammary glands where both luminal and myoepithelial cells are clearly visible. These include the normal tissue from case 159, where myoepithelial cells form a nearly continuous layer surrounding the luminal cells (159N in Figure 1A), and case 402421, where myoepithelial cells are not as prominent but are still noticeable (402421N in Figure 1B). Interestingly, in these normal glands, active modifications H4-acetylation and H3K4me3 and repressive modification H3K9me3 are both significantly depleted in myoepithelial cells (Figure 4A), with the intensity reduced by half in most cases (Figure 4B), when compared to luminal cells.

**In canine complex carcinomas, active modification H4-acetylation is abnormally enriched while repressive modification H3K9me3 is aberrantly depleted.**

Compared to normal mammary glands and simple carcinomas, complex carcinomas harbor significantly more myoepithelial cells (Figure 1). Yet unlike normal mammary glands (Figure 4), both myoepithelial and luminal cells in complex carcinomas were found to be equally enriched with active modifications (Figures 5A-5H). This is especially so for H4-acetylation, with the intensity being equal or stronger than luminal

cells in normal mammary glands and in simple carcinomas. The repressive modification H3K9me3, to the contrary, becomes significantly more depleted in both cell types in complex carcinomas (Figures 5I-5L).

**Redox genes are upregulated in canine ER+ carcinomas, while cell cycle and DNA repair genes are upregulated in canine ER- carcinomas.**

RNA-seq analyses also revealed a clear difference between canine ER+ and ER- tumors (Figure 6A), with most ER+ tumors being complex carcinomas while most ER- tumors being simple carcinomas. Among 1,350 differentially expressed genes at FDR $\leq$ 0.2, approximately half are upregulated in ER+ carcinomas and are significantly enriched in redox functions (Figure 6B). These genes encode ~25 dehydrogenases or oxidases, and 32 gene products are associated with mitochondria including the electron transport chain. Another half of the 1,350 differentially expressed genes are upregulated in ER- carcinomas, among which ~118 genes are associated with the cell cycle, e.g., mitosis, spindle, microtubule cytoskeleton, etc. (Figure 6B). Other significant functions comprise DNA repair (38 genes) and protein serine/threonine kinase activity (17 genes). The same overall conclusions were reached at FDR $\leq$ 0.1.

**Canine simple carcinomas and the ER- complex carcinoma cluster closely with basal-like human breast carcinomas in PAM50 classification.**

To directly compare the canine MCs to human breast cancers, we randomly selected 20 human tumors for each subtype among a total of 195 luminal A, 111 luminal B, 53 HER2-enriched and 87 basal-like tumors of the TCGA RNA-seq study (Cancer

Genome Atlas 2012). This, along with all 7 normal-like tumors in TCGA, amount to 87 human carcinomas covering all five intrinsic subtypes. We then performed PAM50 clustering (Prat and Perou 2011) on these 87 human carcinomas together with our 12 canine carcinomas. This analysis was repeated 100 times, ensuring that each TCGA tumor was sampled at least once. Notably, in 82 out of 100 times, all canine simple carcinomas and the ER- complex carcinoma (ID 518) group with the human basal-like tumors. The remaining canine complex carcinomas (all ER+), however, fail to cluster with any specific human subtypes. One clustering example is shown in Figure 6C.

## DISCUSSION

In this study, we performed an initial comprehensive characterization of the genomes, transcriptomes and epigenomes of two major canine MC histologic subtypes. Even with a small sample size (12 cases), the analysis reveals a remarkable molecular heterogeneity of spontaneous canine MCs. It also emphasizes their unique value and raises a number of important questions that could profoundly impact human breast cancer research.

### Canine simple carcinomas, without myoepithelial cell proliferation, harbor extensive genomic aberrations and are molecularly homologous to human breast carcinomas.

Canine simple carcinomas investigated have no myoepithelial cell proliferation and are histologically comparable to human *in situ* or invasive ductal or lobular

carcinomas. Significantly, these canine cancers faithfully recapitulate key molecular features of human breast cancer. First, analogous to their human counterparts (Volik et al. 2003; Naylor et al. 2005; Sjoblom et al. 2006; Banerji et al. 2012; Cancer Genome Atlas 2012; Curtis et al. 2012; Stephens et al. 2012), the genomes of these canine carcinomas harbor extensive genetic lesions including numerous CNAs, fusion gene-creating translocation, equally complex superamplicon and comparable sequence mutation types. The only exception is an inflammatory carcinoma, whose human equivalent (inflammatory breast cancer) is also devoid of CNAs (Curtis et al. 2012). Second, notable human breast cancer genes (Futreal et al. 2004; Samuels et al. 2004; Sjoblom et al. 2006; Banerji et al. 2012; Cancer Genome Atlas 2012; Stephens et al. 2012) are altered in these canine cancers as well. Examples include: 1) amplification and/or overexpression of oncogenes *BRAF*, *MYC, PIK3CA, PIK3R1, CCND3* and *TBX3*; 2) deletion and/or underexpression of tumor suppressors *PTEN, PTPRD* and *CDH1* (Wendt et al. 2011); and 3) mutations of *BRCA1, NF1, MAP3K1* and *RUNX1*. Third, many of the altered pathways are shared between the two species, e.g., cell adhesion, Wnt-signaling, PI3K signaling and DNA repair (Figure 7C) (Cancer Genome Atlas 2012), consistent with other canine MC studies (Rao et al. 2009; Klopfleisch et al. 2010).

These strong molecular homologies make canine simple carcinomas valuable in human breast cancer research. For example, for cancers with large genomic CNAs, we can apply the dog-human comparison strategy for effective driver-passenger discrimination as described (Tang et al. 2013). Critically, as elegantly discussed in several publications (Paoloni and Khanna 2008; Gordon et al. 2009; Rowell et al. 2011),

these canine cancers, which bridge a gap between traditional rodent models and human clinical trials, can significantly speed up new anticancer drug development. For example, for drugs targeting the PI3K pathway (Gordon and Banerji 2013) (Figure 7C), their efficacy, toxicity, dosage and schedule can be more accurately evaluated through clinical trials with canine patients, before entering human clinical trials. This will significantly reduce the cost and accelerate the drug discovery process.

**Can canine simple carcinomas serve as a much-needed spontaneous cancer model of basal-like human breast cancer?**

Canine simple carcinomas cluster with basal-like human tumors with an 82% chance in our PAM50 classification, indicating their closer resemblance to this subtype than other intrinsic subtypes. This may be explained by the observation that none of the canine tumors carrying HER2 amplification or overexpression. Furthermore, many harbor extensive CNAs and are ER- with genes related to DNA repair and cell cycle significantly upregulated, consistent with the basal-like subtype profile (Cancer Genome Atlas 2012). This is especially so considering that the ER- complex carcinoma clusters similarly as well. The only ER+ canine simple carcinoma has a prominent *PTEN* deletion, also a feature of basal-like tumors (Cancer Genome Atlas 2012).

Clearly, studies with a larger sample size are needed to determine if canine simple carcinomas as a whole or even just a subset do indeed closely match the basal-like subtype. If confirmed, these canine cancers could serve as a much-needed spontaneous cancer model. Compared to other subtypes, basal-like cancers are aggressive, have a

poor prognosis and currently lack effective treatments. Canine MC could make significant contributions towards understanding and treating this worst subtype of human breast cancer.

**Canine complex carcinomas, with myoepithelial cell proliferation, appear to originate from epigenomic rather than genomic alterations.**

Complex carcinomas, featuring dual proliferation of luminal and myoepithelial cells, likely originate from epigenomic, rather than genomic, abnormalities (Figure 7). First, their genomes appear normal without CNAs detected and with sequence mutation rates as low as normal tissues. Thus, it is unlikely that these tumors arise from genetic aberrations, unlike simple carcinomas. Meanwhile, complex carcinomas could acquire genomic changes as they progress to later stages, as shown by tumor 518 and another complex carcinoma with pulmonary metastasis (Beck et al. 2013). Second, chromatin modification and transcription regulation stand out as the most enriched functions among genes differentially expressed between simple and complex carcinomas, with numerous chromatin-modification genes downregulated in complex carcinomas. Importantly, complex carcinomas are aberrantly enriched with active histone modification H4-acetylation while abnormally depleted with repressive modification H3K9me3. Thus, it is possible that the epigenomes of complex carcinomas are altered, turning on genes that normally should be silenced to promote tumorigenesis. Obviously, more studies are needed to confirm this possibility and to understand the underlying mechanisms.

Myoepithelial cell proliferation is rare in human breast cancer (Hayes 2011; Tan and Ellis 2013). As a result, myoepithelial cells receive far less attention than luminal cells and are poorly understood (Lakhani and O'Hare 2001; Adriance et al. 2005; Sopel 2010; Moumen et al. 2011). However, the few laboratories that study myoepithelial cells have noted their importance. For example, myoepithelial cells are thought to be a part of the mammary stem cell niche, mediate the cross-talk between luminal cells and extracellular matrix, contribute to the maintenance of the apicobasal polarity of luminal cells, and serve as a tumor suppressor (Gudjonsson et al. 2005; Polyak and Hu 2005; Hu et al. 2008). Canine MC, where myoepithelial cell proliferation is much more common, provides an ideal system to address such functions and to better understand the 2$^{nd}$ major cell lineage of the mammary gland (e.g., by answering questions such as whether a prolonged luteal phase promotes myoepithelial cell proliferation).

**Do canine complex carcinomas derive from mammary gland stem cells or luminal/myoepithelial common progenitors?**

Several observations indicate the possibility that complex carcinomas arise from mammary gland stem cells or luminal/myoepithelial common progenitors (Figure 7A). First, one of these tumors (ID 518) expresses the pluripotency marker *SOX2*, indicating stem cell property. Second, unlike normal mammary glands which present a clearly different epigenomic landscape between luminal and myoepithelial cells, no such difference was observed in complex carcinomas. This indicates that proliferating luminal and myoepithelial cells in complex carcinomas may have derived from altered common progenitors. Third, compared to simple carcinomas and normal mammary tissues,

glucose metabolic genes are upregulated in complex carcinomas, consistent with this stem cell or progenitor origin theory. For simple carcinomas, we hypothesize that they originate from either luminal progenitors, because of their close resemblance to the basal-like subtype which is reported to have a luminal progenitor origin (Molyneux et al. 2010), or differentiated luminal cells because of luminal cell properties (see case 159 in Figure 1). Of course, further studies with a larger sample size are needed to test these hypotheses.

In summary, we performed the first comprehensive characterization of the genomes, transcriptomes and epigenomes of canine simple carcinomas and complex carcinomas, two major histologic subtypes of canine MC. The analysis reveals that canine simple carcinomas, which have no myoepithelial cell proliferation and histologically match typical human breast carcinomas, faithfully recapitulate many molecular features of human breast cancer. Notably, canine simple carcinomas closely cluster with basal-like human breast tumors in PAM50 classification, and thus could serve as a much-needed spontaneous cancer model for the basal-like subtype. Canine complex carcinomas are characterized with dual proliferation of luminal and myoepithelial cells, which is rare in human breast cancer. Our analysis indicates that these canine cancers may arise from epigenomic rather genomic alterations. Canine complex carcinomas hence provide a unique system to investigate the roles of myoepithelial cells, the 2$^{nd}$ major cell lineage of the mammary gland, in normal developmental and pathogenic processes.

## MATERIALS AND METHODS

### Canine Mammary Tissue Samples

Fresh-frozen (FF) and formalin-fixed paraffin-embedded (FFPE) normal and tumor tissue samples of spontaneous canine MC were obtained from the University of California-Davis School of Veterinary Medicine and the Animal Cancer Tissue Repository of the Colorado State University. Samples were collected from client-owned dogs that develop the disease spontaneously, under the guidelines of the Institutional Animal Care and Use Committee and with owner informed consent.

### Immunohistochemical (IHC) Analyses

IHC experiments were performed following standard protocols with 5 μm FFPE sections. Primary antibodies were used as described (Bryson et al. 2013), including those against smooth muscle myosin heavy chain clone ID8 (MAB3568), acetyl-H4 (06-866), H3K9me3 (07-442) and H3-K4/K9-me3 (06-866), all from Millipore; H3K4me3 (Abcam, ab8580), estrogen receptor alpha clone E115 (Abcam, ab32063); and E-cadherin (R&D Systems, AF648). Alexa Fluor® 488, 647 or 594 conjugated secondary antibodies are from Jackson ImmunoResearch. Images were taken with a Zeiss LSM 710 confocal microscope.

### Tissue Dissection, DNA and RNA Extraction, and PCR Analyses

Cryosectioning of FF tissues, H&E staining and cryomicrodissection were performed as described (Tang et al. 2010) to enrich tumor cells for tumor samples and mammary gland epithelial cells for normal samples. Genomic DNA and RNA were then

extracted from the dissected tissues using the DNeasy Blood & Tissue Kit (cat. no. 69504), RNeasy Plus Mini Kit (cat. no. 74134) or AllPrep DNA/RNA Mini Kit (cat. no. 80204) from QIAGEN. Only samples with a 260/280 ratio of ~1.8 (DNA) or ~2.0 (RNA) and showing no degradation and other contaminations on the agarose gels were subjected to further analyses. The synthesis of cDNA, primer design, and PCR or qPCR with genomic DNA or cDNA samples were conducted as described (Youmans et al. 2012).

**aCGH Analyses**

aCGH experiments were conducted at the Florida State University Microarray Facility, with 385K canine CGH array chips from Roche NimbleGen Systems, Inc. CNAs were identified as described (Tang et al. 2010).

**Paired-end WGS, WES and RNA-seq**

All three types of sequencing were conducted using the Illumina platform, following the protocols from the manufacture. Paired-end WGS of >12X sequence coverage was performed in collaboration with the Emory Genome Center (50bp or 100bp paired-end sequencing of ~200bp fragments) or the BGI-America (90bp paired-end sequencing of ~500bp fragments). WES was conducted in collaboration with the Hudsonalpha Institute for Biotechnology. First, exome-capturing was achieved by using a solution-based SureSelect kit from Agilent, covering 50Mb canine exons and adjacent regions. Then, paired-end sequences of 50bp of ~200bp fragments were generated from the captured targets to reach the coverage of 134-245X. RNA-seq was performed at Hudsonalpha, yielding 42 to 94 million paired-end sequence reads of 50bp per sample.

## Sequence Data Analyses

Briefly, WGS, WES and RNA-seq sequence reads were aligned to the dog reference genome (Lindblad-Toh et al. 2005). Then, uniquely mapped WES reads were used to detect base substitutions and small indels, and significantly mutated genes were identified as described (Sjoblom et al. 2006). Uniquely mapped WGS read pairs were used to identify somatic translocations and chimeric fusion genes. Uniquely mapped RNA-seq read pairs were used to quantify each gene's expression level, as well as to detect chimeric fusion transcripts and sequence mutations.

## Data Access

Sequence data have been submitted to the NCBI SRA database with accession numbers SRP023115, SRP023472 and SRP024250. aCGH data have been submitted to the GEO database with accession number GSE54535.

## Acknowledgement

We thank Drs. Timothy D. Read, Shawn Levy, Steven H. Miller, and their respective teams, and the BGI, for their outstanding sequencing or aCGH work; Ms. Teri Guerrero, Ms. Irene Mok and Dr. Susan E Lana for their help on sample collection; Dr. John M. Rosenfeld of Millipore for providing antibodies; Drs. Nancy Manley and Jie Li for their help on IHC and H&E analyses; and Drs. Dong M Shin, J David Puett, Charles M Perou and Malcolm Hayes for their help on the study. Confocal imaging was performed at the UGA Biomedical Microscopy Core.

## FIGURE LEGENDS

**Figure 1. Myoepithelial cell proliferation is absent in canine simple carcinomas but prominent in canine complex carcinomas**.

**A and B**, representative images of immunostaining with the myoepithelial marker SMHC and the luminal marker E-cadherin (E-cad) of normal (N) and tumor (T) tissues of two simple carcinoma cases (A), one *in situ* (ID 159) and the other invasive (ID 401188), and two complex carcinomas (B). Top panel shows the enlarged view of the areas indicated below. Red arrows point to luminal cells, while yellow arrows point to myoepithelial cells. Scale bar = 100 μm.

**C and D**, H&E staining of the same tissues.

**Figure 2. Large scale genomic aberrations are frequent in canine simple carcinomas but rare in canine complex carcinomas**.

**A**, CNAs found in four complex carcinomas (labeled), one half complex and half simple carcinoma (ID 32510) and 7 simple carcinomas (which include the inflammatory tumor 115 and the six tumors at the 2$^{nd}$ panel) by aCGH. The images were drawn as described (Tang et al. 2010), with each line representing a canine chromosome and vertical lines **above**/**below** the chromosome indicating amplifications/deletions, respectively. Notable amplified/deleted genes are shown.

**B**, the total numbers of amplified (shaded bars) and deleted (empty bars) genes of each carcinoma shown in A.

**C**, two >4Mb amplicons discovered in simple carcinoma 76 in A, by both WGS and aCGH. The X-axis indicates chromosomal coordinates in Mb, while the Y-axis indicates

the mapped read pair density (MPD) values of WGS or the tumor against normal $\log_2$-ratios of aCGH.

**D**, the proposed mechanism for superamplicon formation. Prior sequence amplifications led to two translocations (represented by the dashed lines), resulting in a circle which was further amplified. The numbers indicate the chromosomal coordinates in bp.

**E**, a fusion gene created by the 2$^{nd}$ translocation shown in D. The translocation occurred in the intron of both genes as indicated (exons are represented by the vertical bars). An in-frame fusion transcript then emerged via splicing.

**F**, the A20-type domain of ZFAND3 and the glucoamylase domain of MGAM are preserved in the fusion protein.

**Figure 3. Coding sequence mutations are frequent in canine simple carcinomas; chromatin-modification genes are downregulated in canine complex carcinomas**.

**A**, the fractions (the Y-axis) of somatic base substitution types of simple carcinomas (IDs indicated by the X-axis) detected by WES. The total number of significantly mutated genes in each tumor is also shown, and tumor 5 has many DNA repair genes mutated.

**B**, synonymous (green dots) and non-synonymous substitutions (yellow dots), and a nonsense mutation (red star) uncovered in the *USH2A* gene in tumor 5.

**C**, the base substitution (compared to the dog reference genome) rates of the three sample types in coding regions with 30-300X RNA-seq read coverage. The p-values were calculated by t-tests.

**D**, the heatmap of 751 genes differentially expressed at FDR $\leq 0.2$ between simple and complex carcinomas (red: upregulation; green: downregulation). The right panel

illustrates the enriched functions of each gene cluster indicated, and the 35 chromatin modifiers downregulated in complex carcinomas are indicated .

**Figure 4. In canine normal mammary glands, active and repressive histone modifications are both depleted in myoepithelial cells when compared to luminal cells.**

**A**, representative IHC images of active modifications acetyl-H4 and H3K4me3, repressive modification H3K9me3, and modification H3-K4/K9-me3 (positive only when H3K4me3 and H3K9me3 are present simultaneously). Yellow arrows point to myoepithelial cells, while red arrows point to luminal cells. Scale bar = 100 μm.

**B**, the intensity of each histone modification was measured from ≥ 10 individual cells of each type from different regions across the tissue section. The p-values were calculated by Wilcoxon tests.

**Figure 5. In canine complex carcinomas, active histone modification H4-acetylation is enriched while repressive modification H3K9me3 is depleted in both luminal and myoepithelial cells.**

**A-C**, representative IHC images of acetyl-H4 of simple carcinomas (A), complex carcinomas (B), and normal mammary glands (C). The merged (top) and split images (Acetyl: acetyl-H4) are shown. Scale bar = 100 μm.

**D**, the immunofluorescence staining intensity of acetyl-H4 determined from at least three different areas of the first split image, labeled as "Acetyl", in A-C. The pairwise comparison p-values were calculated by Wilcoxon tests.

**E-H** for H3K4me3 (K4) and **I-L** for H3K9me3 (K9) are presented in the same way as A-D. Unlike the 159N sample of Figure 4A, the normal mammary glands in C, G and K consist of mostly luminal cells, with myoepithelial cells either absent (401188N) or very few (402421N).

**Figure 6. Subtype analysis reveals homology between canine MCs and human breast cancers.**

**A**, examples of ER+ (ID 401188) and ER- (ID 341400) canine carcinomas determined by IHC. Scale bar = 100 μm.

**B**, genes (1,350 total) differentially expressed between ER+ and ER- canine carcinomas at FDR ≤ 0.2. The tumor IDs are indicated. Tumor 401188 (purple) is ER+ but a simple carcinomas, while tumor 518 (orange) is ER- but a complex carcinoma. The right panel illustrates the significantly enriched functions of each gene cluster indicated.

**C**, Canine simple carcinomas and the ER- complex carcinoma cluster with the basal-like human breast carcinomas in PAM50 classification. The heatmap represents a clustering example of 12 canine tumors and 87 TCGA human tumors (see text). The composition of each cluster specified at the top of the heatmap is explained at the right.

**Figure 7. Canine complex carcinomas possibly arise from epigenomic alterations, whereas canine simple carcinomas likely originate from genomic aberrations.**

**A**, the proposed carcinogenic mechanism. The mammary gland development hierarchy is modified from a publication (Lim et al. 2009).

**B**, epigenomic alterations in complex carcinomas, with histone modifications enriched (darker shading) or depleted (lighter shading).

**C**, genomic alterations in simple carcinomas, with notable gene and pathway alterations (activation: darker shading; inactivation: lighter shading) indicated in the respective tumors (e.g., PTEN deletion in tumor 401188).

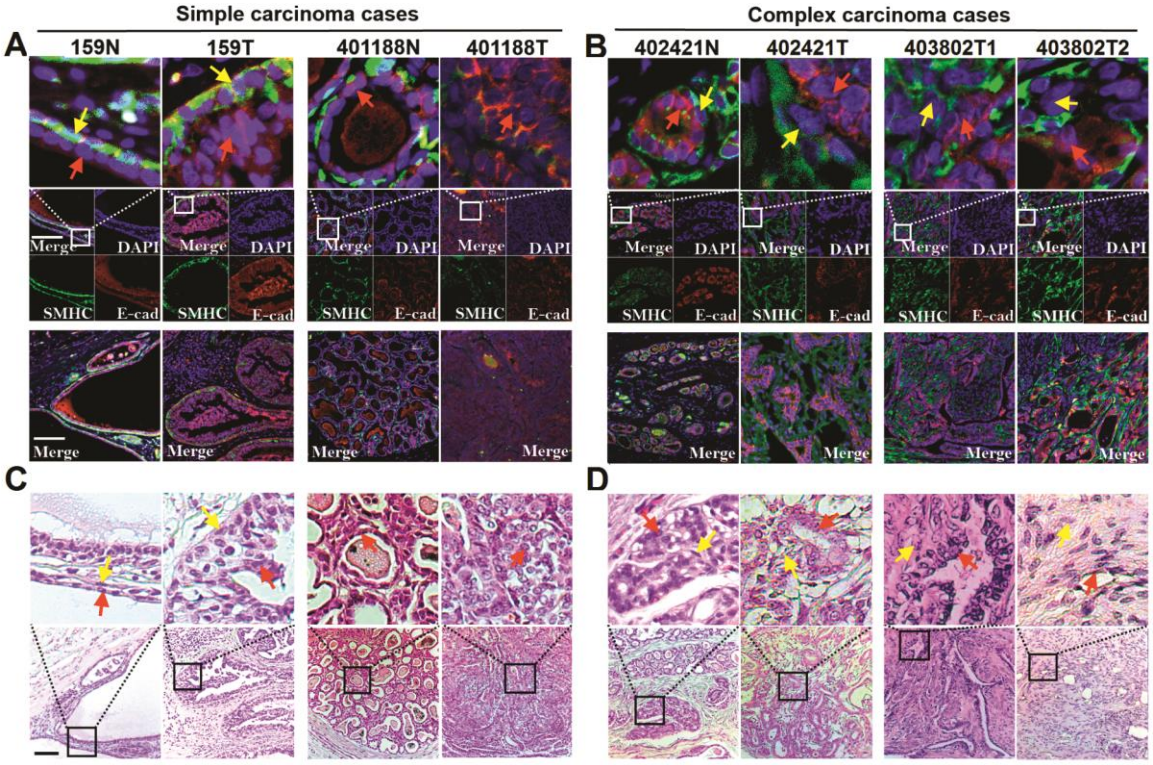

Simple carcinoma cases
Complex carcinoma cases
A
159N
159T
401188N
401188T
B
402421N
402421T
403802T1
403802T2
Merge
DAPI
SMHC
E-cad
C
D

Figure 1

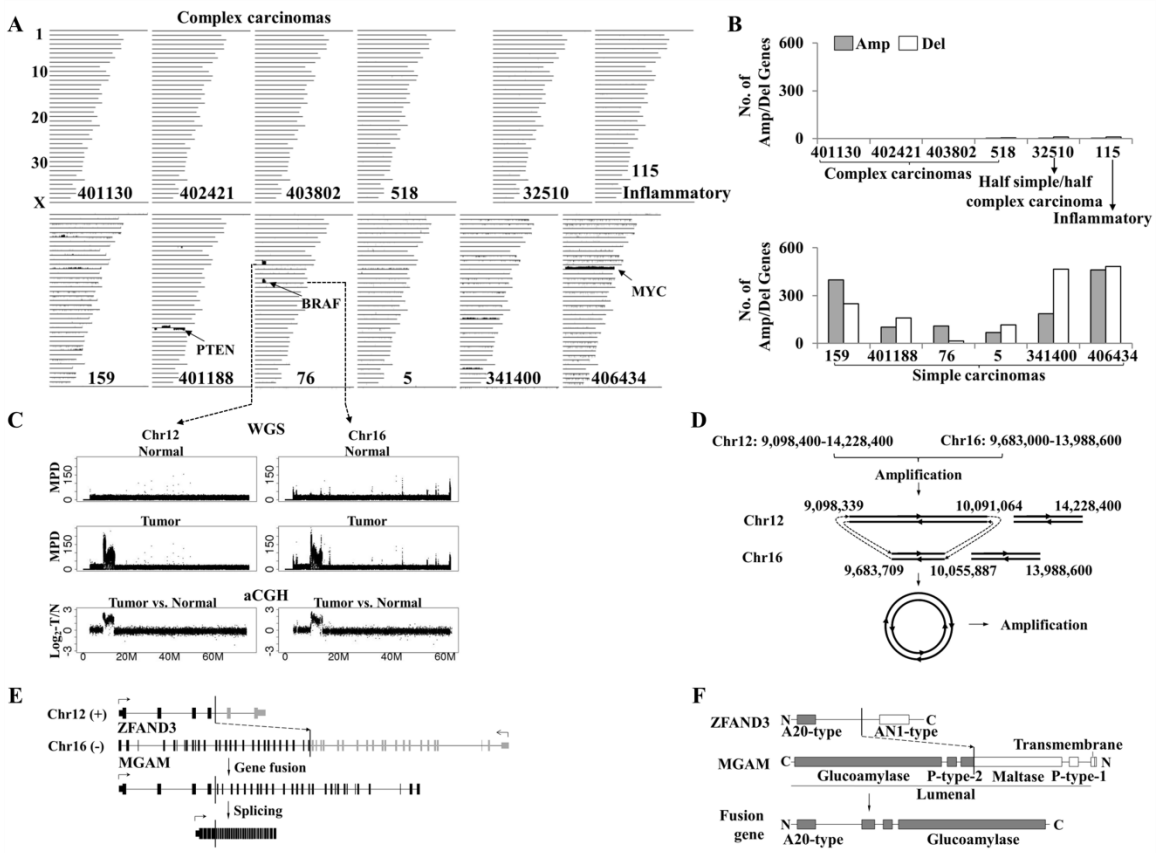

A
Complex carcinomas
1
10
20
30
X
401130
402421
403802
518
32510
115
Inflammatory
BRAF
MYC
PTEN
159
401188
76
5
341400
406434
B
Amp
Del
No. of Amp/Del Genes
600
300
0
401130 402421 403802 518 32510 115
Complex carcinomas
Half simple/half complex carcinoma
Inflammatory
159 401188 76 5 341400 406434
Simple carcinomas
C
WGS
Chr12 Normal
Chr16 Normal
Tumor
MPD
aCGH
Tumor vs. Normal
Log2-T/N
0 20M 40M 60M
D
Chr12: 9,098,400-14,228,400
Chr16: 9,683,000-13,988,600
Amplification
9,098,339
10,091,064
14,228,400
Chr12
Chr16
9,683,709
10,055,887
13,988,600
E
Chr12 (+)
ZFAND3
Chr16 (-)
MGAM
Gene fusion
Splicing
F
ZFAND3
A20-type
AN1-type
Transmembrane
MGAM
Glucoamylase
P-type-2
Maltase
P-type-1
Lumenal
Fusion gene

**Figure 2**

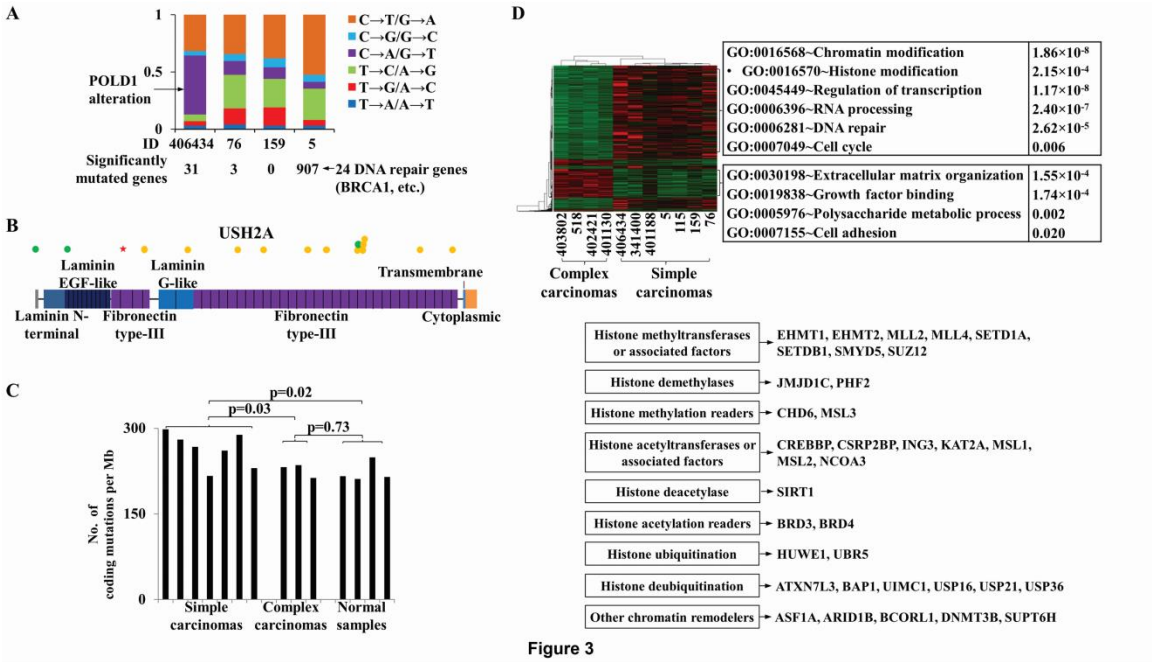

A
1
0.5
0
POLD1 alteration
C→T/G→A
C→G/G→C
C→A/G→T
T→C/A→G
T→G/A→C
T→A/A→T
ID 406434 76 159 5
Significantly mutated genes 31 3 0 907
24 DNA repair genes (BRCA1, etc.)
B
USH2A
Laminin EGF-like
Laminin G-like
Transmembrane
Laminin N-terminal
Fibronectin type-III
Fibronectin type-III
Cytoplasmic
C
p=0.02
p=0.03
p=0.73
300
150
0
No. of coding mutations per Mb
Simple carcinomas
Complex carcinomas
Normal samples
D
403802
518
402421
401130
406434
341400
401188
5
115
159
76
Complex carcinomas
Simple carcinomas
GO:0016568~Chromatin modification 1.86×10-8
• GO:0016570~Histone modification 2.15×10-4
GO:0045449~Regulation of transcription 1.17×10-8
GO:0006396~RNA processing 2.40×10-7
GO:0006281~DNA repair 2.62×10-5
GO:0007049~Cell cycle 0.006
GO:0030198~Extracellular matrix organization 1.55×10-4
GO:0019838~Growth factor binding 1.74×10-4
GO:0005976~Polysaccharide metabolic process 0.002
GO:0007155~Cell adhesion 0.020
Histone methyltransferases or associated factors → EHMT1, EHMT2, MLL2, MLL4, SETD1A, SETDB1, SMYD5, SUZ12
Histone demethylases → JMJD1C, PHF2
Histone methylation readers → CHD6, MSL3
Histone acetyltransferases or associated factors → CREBBP, CSRP2BP, ING3, KAT2A, MSL1, MSL2, NCOA3
Histone deacetylase → SIRT1
Histone acetylation readers → BRD3, BRD4
Histone ubiquitination → HUWE1, UBR5
Histone deubiquitination → ATXN7L3, BAP1, UIMC1, USP16, USP21, USP36
Other chromatin remodelers → ASF1A, ARID1B, BCORL1, DNMT3B, SUPT6H

**Figure 3**

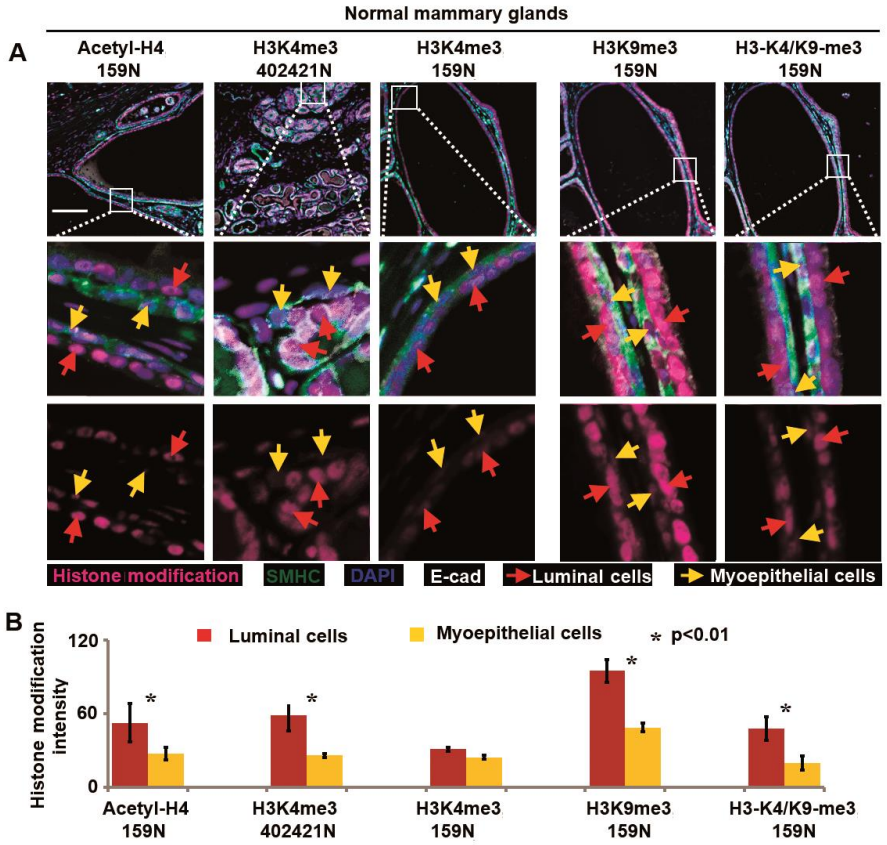

Normal mammary glands
A
Acetyl-H4 159N
H3K4me3 402421N
H3K4me3 159N
H3K9me3 159N
H3-K4/K9-me3 159N
Histone modification
SMHC
DAPI
E-cad
Luminal cells
Myoepithelial cells
B
Histone modification intensity
120
60
0
Luminal cells
Myoepithelial cells
* p<0.01
Acetyl-H4 159N
H3K4me3 402421N
H3K4me3 159N
H3K9me3 159N
H3-K4/K9-me3 159N

Figure 4

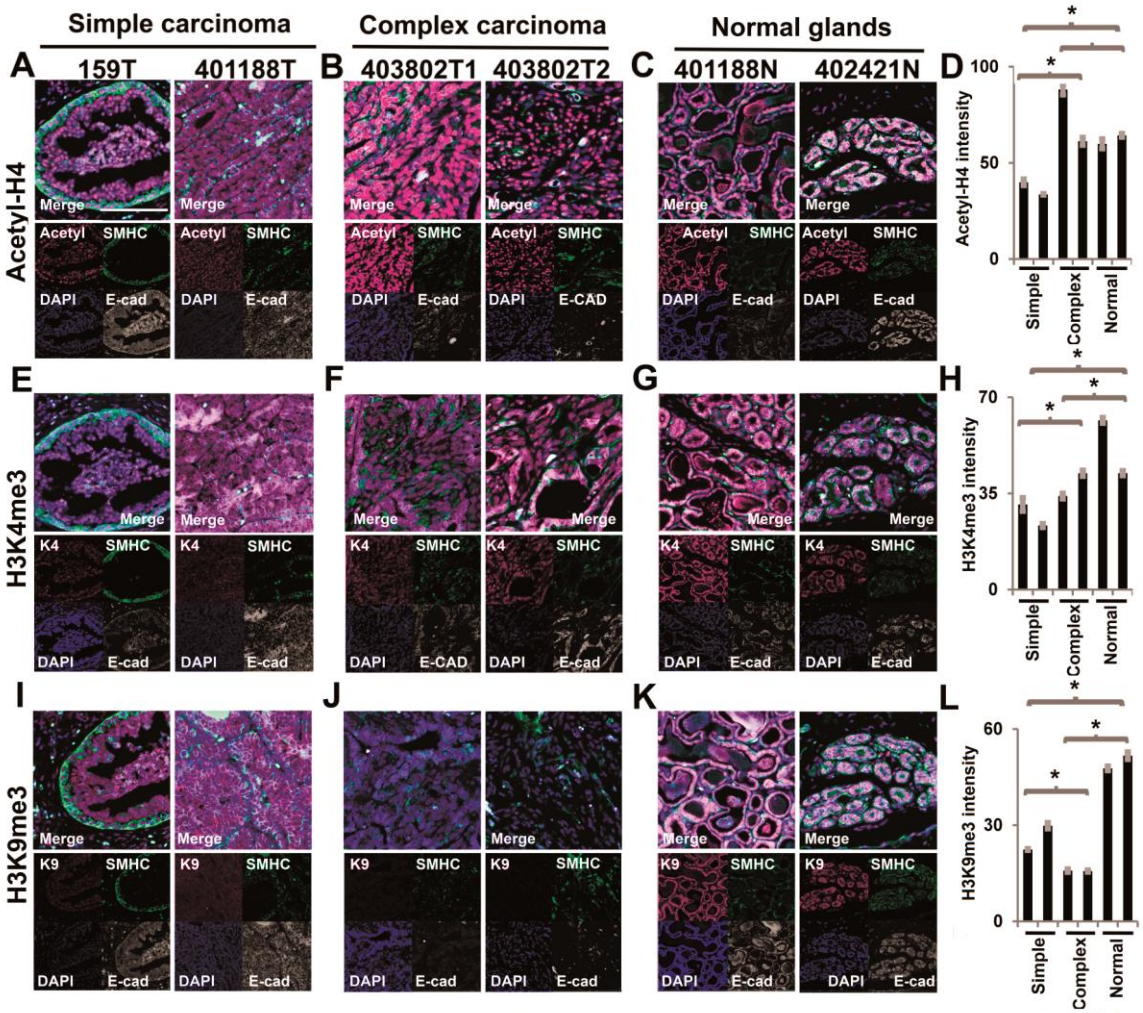

Simple carcinoma
Complex carcinoma
Normal glands
A
159T
401188T
B
403802T1
403802T2
C
401188N
402421N
D
Acetyl-H4
H3K4me3
H3K9me3
Merge
Acetyl
SMHC
DAPI
E-cad
E-CAD
K4
K9
Acetyl-H4 intensity
H3K4me3 intensity
H3K9me3 intensity
Simple
Complex
Normal
100
50
0
70
35
60
30
* p<0.01

Figure 5

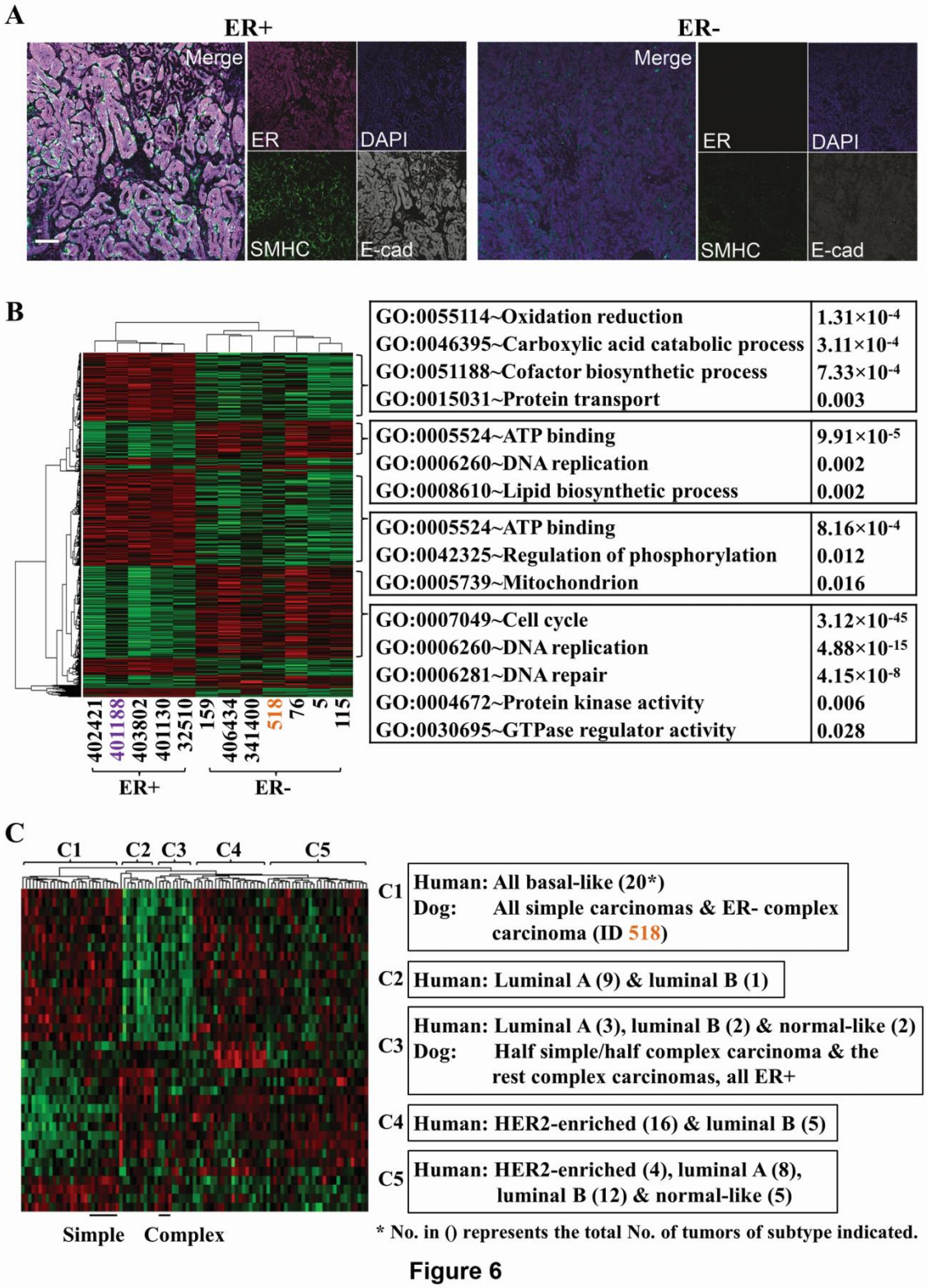

A
ER+
ER-
Merge
ER
DAPI
SMHC
E-cad
B
GO:0055114~Oxidation reduction 1.31×10-4
GO:0046395~Carboxylic acid catabolic process 3.11×10-4
GO:0051188~Cofactor biosynthetic process 7.33×10-4
GO:0015031~Protein transport 0.003
GO:0005524~ATP binding 9.91×10-5
GO:0006260~DNA replication 0.002
GO:0008610~Lipid biosynthetic process 0.002
GO:0005524~ATP binding 8.16×10-4
GO:0042325~Regulation of phosphorylation 0.012
GO:0005739~Mitochondrion 0.016
GO:0007049~Cell cycle 3.12×10-45
GO:0006260~DNA replication 4.88×10-15
GO:0006281~DNA repair 4.15×10-8
GO:0004672~Protein kinase activity 0.006
GO:0030695~GTPase regulator activity 0.028
402421
401188
403802
401130
32510
159
406434
341400
518
76
5
115
ER+
ER-
C
C1
C2
C3
C4
C5
Simple
Complex
C1 Human: All basal-like (20*)
Dog: All simple carcinomas & ER- complex carcinoma (ID 518)
C2 Human: Luminal A (9) & luminal B (1)
C3 Human: Luminal A (3), luminal B (2) & normal-like (2)
Dog: Half simple/half complex carcinoma & the rest complex carcinomas, all ER+
C4 Human: HER2-enriched (16) & luminal B (5)
C5 Human: HER2-enriched (4), luminal A (8), luminal B (12) & normal-like (5)
* No. in () represents the total No. of tumors of subtype indicated.

Figure 6

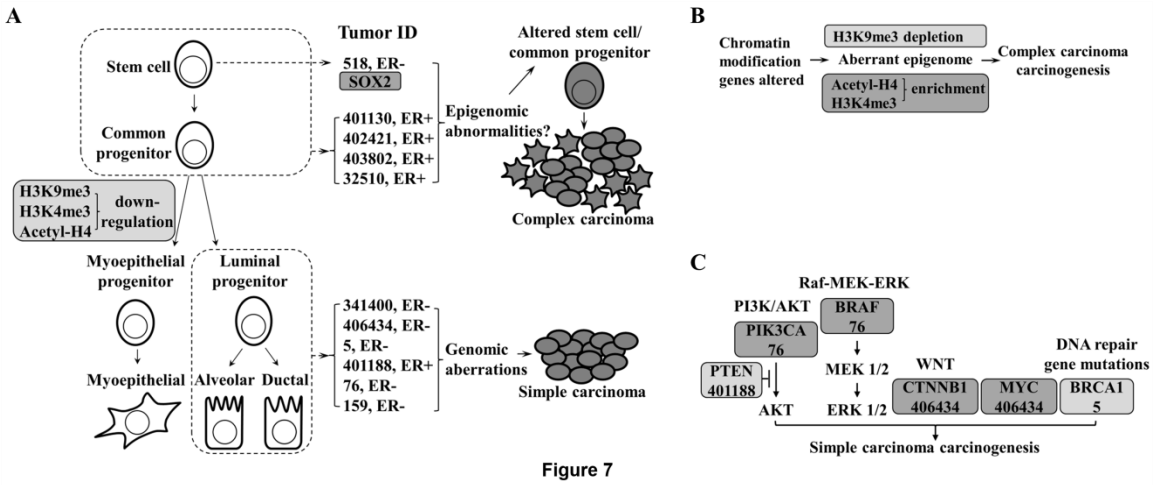

A
Tumor ID
Altered stem cell/
common progenitor
Stem cell
518, ER-
SOX2
Common
progenitor
401130, ER+
402421, ER+
403802, ER+
32510, ER+
Epigenomic
abnormalities?
Complex carcinoma
H3K9me3
H3K4me3
Acetyl-H4
down-
regulation
Myoepithelial
progenitor
Luminal
progenitor
341400, ER-
406434, ER-
5, ER-
401188, ER+
76, ER-
159, ER-
Genomic
aberrations
Simple carcinoma
Myoepithelial
Alveolar
Ductal
B
Chromatin
modification
genes altered
H3K9me3 depletion
Aberrant epigenome
Complex carcinoma
carcinogenesis
Acetyl-H4
H3K4me3
enrichment
C
Raf-MEK-ERK
PI3K/AKT
BRAF
76
PIK3CA
76
MEK 1/2
WNT
DNA repair
gene mutations
PTEN
401188
CTNNB1
406434
MYC
406434
BRCA1
5
AKT
ERK 1/2
Simple carcinoma carcinogenesis

Figure 7

## REFERENCES


Adriance MC, Inman JL, Petersen OW, Bissell MJ. 2005. Myoepithelial cells: good fences make good neighbors. *Breast cancer research : BCR* **7**(5): 190-197.

Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, Lawrence MS, Sivachenko AY, Sougnez C, Zou L et al. 2012. Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature* **486**(7403): 405-409.

Beck J, Hennecke S, Bornemann-Kolatzki K, Urnovitz HB, Neumann S, Strobel P, Kaup FJ, Brenig B, Schutz E. 2013. Genome aberrations in canine mammary carcinomas and their detection in cell-free plasma DNA. *PloS one* **8**(9): e75485.

Bryson JL, Griffith AV, Hughes B, 3rd, Saito F, Takahama Y, Richie ER, Manley NR. 2013. Cell-autonomous defects in thymic epithelial cells disrupt endothelial-perivascular cell interactions in the mouse thymus. *PloS one* **8**(6): e65196.

Cancer Genome Atlas N. 2012. Comprehensive molecular portraits of human breast tumours. *Nature* **490**(7418): 61-70.

Costello M, Pugh TJ, Fennell TJ, Stewart C, Lichtenstein L, Meldrim JC, Fostel JL, Friedrich DC, Perrin D, Dionne D et al. 2013. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic acids research*.

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y et al. 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**(7403): 346-352.

Dawson MA, Kouzarides T. 2012. Cancer epigenetics: from mechanism to therapy. *Cell* **150**(1): 12-27.

Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, Bonaldi T, Haydon C, Ropero S, Petrie K et al. 2005. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nature genetics* **37**(4): 391-400.

Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N, Stratton MR. 2004. A census of human cancer genes. *Nature reviews Cancer* **4**(3): 177-183.

Goldschmidt M, Pena L, Rasotto R, Zappulli V. 2011. Classification and grading of canine mammary tumors. *Veterinary pathology* **48**(1): 117-131.

Gordon I, Paoloni M, Mazcko C, Khanna C. 2009. The Comparative Oncology Trials Consortium: using spontaneously occurring cancers in dogs to inform the cancer drug development pathway. *PLoS medicine* **6**(10): e1000161.

Gordon V, Banerji S. 2013. Molecular pathways: PI3K pathway targets in triple-negative breast cancers. *Clinical cancer research : an official journal of the American Association for Cancer Research* **19**(14): 3738-3744.

Gudjonsson T, Adriance MC, Sternlicht MD, Petersen OW, Bissell MJ. 2005. Myoepithelial cells: their origin and function in breast morphogenesis and neoplasia. *Journal of mammary gland biology and neoplasia* **10**(3): 261-272.

Hayes MM. 2011. Adenomyoepithelioma of the breast: a review stressing its propensity for malignant transformation. *Journal of clinical pathology* **64**(6): 477-484.

Hu M, Yao J, Carroll DK, Weremowicz S, Chen H, Carrasco D, Richardson A, Violette S, Nikolskaya T, Nikolsky Y et al. 2008. Regulation of in situ to invasive breast carcinoma transition. *Cancer cell* **13**(5): 394-406.

Ji X, Zhao S. 2008. DA and Xiao-two giant and composite LTR-retrotransposon-like elements identified in the human genome. *Genomics* **91**(3): 249-258.

Klopfleisch R, Lenze D, Hummel M, Gruber AD. 2010. Metastatic canine mammary carcinomas can be identified by a gene expression profile that partly overlaps with human breast cancer profiles. *BMC cancer* **10**: 618.

Lademann U, Kallunki T, Jaattela M. 2001. A20 zinc finger protein inhibits TNF-induced apoptosis and stress response early in the signaling cascades and independently of binding to TRAF2 or 14-3-3 proteins. *Cell death and differentiation* **8**(3): 265-272.

Lakhani SR, O'Hare MJ. 2001. The mammary myoepithelial cell--Cinderella or ugly sister? *Breast cancer research : BCR* **3**(1): 1-4.

Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, Asselin-Labat ML, Gyorki DE, Ward T, Partanen A et al. 2009. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nature medicine* **15**(8): 907-913.

Lindblad-Toh K Wade CM Mikkelsen TS Karlsson EK Jaffe DB Kamal M Clamp M Chang JL Kulbokas EJ, 3rd Zody MC et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**(7069): 803-819.

Meuten DJ. 2002. *Tumors in domestic animals*. Iowa State University Press, Ames, Iowa.

Molyneux G, Geyer FC, Magnay FA, McCarthy A, Kendrick H, Natrajan R, Mackay A, Grigoriadis A, Tutt A, Ashworth A et al. 2010. BRCA1 basal-like breast cancers

originate from luminal epithelial progenitors and not from basal stem cells. *Cell stem cell* **7**(3): 403-417.

Moumen M, Chiche A, Cagnet S, Petit V, Raymond K, Faraldo MM, Deugnier MA, Glukhova MA. 2011. The mammary myoepithelial cell. *The International journal of developmental biology* **55**(7-9): 763-771.

Nasir L, Devlin P, McKevitt T, Rutteman G, Argyle DJ. 2001. Telomere lengths and telomerase activity in dog tissues: a potential model system to study human telomere and telomerase biology. *Neoplasia* **3**(4): 351-359.

Naylor TL, Greshock J, Wang Y, Colligon T, Yu QC, Clemmer V, Zaks TZ, Weber BL. 2005. High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization. *Breast cancer research : BCR* **7**(6): R1186-1198.

Palles C, Cazier JB, Howarth KM, Domingo E, Jones AM, Broderick P, Kemp Z, Spain SL, Almeida EG, Salguero I et al. 2012. Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nature genetics* **45**(2): 136-144.

Paoloni M, Khanna C. 2008. Translation of new cancer treatments from pet dogs to humans. *Nature reviews Cancer* **8**(2): 147-156.

Parker HG, Shearin AL, Ostrander EA. 2010. Man's best friend becomes biology's best in show: genome analyses in the domestic dog. *Annual review of genetics* **44**: 309-336.

Paw Owski KM, Maciejewski H, Dolka I, Mol JA, Motyl T, Krol M. 2013. Gene expression profiles in canine mammary carcinomas of various grades of malignancy. *BMC veterinary research* **9**(1): 78.

Polyak K, Hu M. 2005. Do myoepithelial cells hold the key for breast tumor progression? *Journal of mammary gland biology and neoplasia* **10**(3): 231-247.

Prat A, Perou CM. 2011. Deconstructing the molecular portraits of breast cancer. *Molecular oncology* **5**(1): 5-23.

Rangarajan A, Weinberg RA. 2003. Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nature reviews Cancer* **3**(12): 952-959.

Rao NA, van Wolferen ME, Gracanin A, Bhatti SF, Krol M, Holstege FC, Mol JA. 2009. Gene expression profiles of progestin-induced canine mammary hyperplasia and spontaneous mammary tumors. *Journal of physiology and pharmacology : an official journal of the Polish Physiological Society* **60 Suppl 1**: 73-84.

Rowell JL, McCarthy DO, Alvarez CE. 2011. Dog models of naturally occurring cancer. *Trends in molecular medicine* **17**(7): 380-388.

Samuels Y, Wang ZH, Bardelli A, Silliman N, Ptak J, Szabo S, Yan H, Gazdar A, Powell DM, Riggins GJ et al. 2004. High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304**(5670): 554-554.

Sauer J, Sigurskjold BW, Christensen U, Frandsen TP, Mirgorodskaya E, Harrison M, Roepstorff P, Svensson B. 2000. Glucoamylase: structure/function relationships, and protein engineering. *Biochimica et biophysica acta* **1543**(2): 275-293.

Siegel R, Naishadham D, Jemal A. 2012. Cancer statistics, 2012. *CA: a cancer journal for clinicians* **62**(1): 10-29.

Sjoblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N et al. 2006. The consensus coding sequences of human breast and colorectal cancers. *Science* **314**(5797): 268-274.

Sleeckx N, de Rooster H, Veldhuis Kroeze EJ, Van Ginneken C, Van Brantegem L. 2011. Canine mammary tumours, an overview. *Reproduction in domestic animals = Zuchthygiene* **46**(6): 1112-1131.

Sopel M. 2010. The myoepithelial cell: its role in normal mammary glands and breast cancer. *Folia morphologica* **69**(1): 1-14.

Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C, Wedge DC, Nik-Zainal S, Martin S, Varela I, Bignell GR et al. 2012. The landscape of cancer genes and mutational processes in breast cancer. *Nature* **486**(7403): 400-404.

Tan PH, Ellis IO. 2013. Myoepithelial and epithelial-myoepithelial, mesenchymal and fibroepithelial breast lesions: updates from the WHO Classification of Tumours of the Breast 2012. *Journal of clinical pathology*.

Tang J, Le S, Sun L, Yan X, Zhang M, Macleod J, Leroy B, Northrup N, Ellis A, Yeatman TJ et al. 2010. Copy number abnormalities in sporadic canine colorectal cancers. *Genome research* **20**(3): 341-350.

Tang J, Li Y, Lyon K, Camps J, Dalton S, Ried T, Zhao S. 2013. Cancer driver-passenger distinction via sporadic human and dog cancer comparison: a proof-of-principle study with colorectal cancer. *Oncogene*.

Volik S, Zhao S, Chin K, Brebner JH, Herndon DR, Tao Q, Kowbel D, Huang G, Lapuk A, Kuo WL et al. 2003. End-sequence profiling: sequence-based analysis of aberrant genomes. *Proceedings of the National Academy of Sciences of the United States of America* **100**(13): 7696-7701.

Wendt MK, Taylor MA, Schiemann BJ, Schiemann WP. 2011. Down-regulation of epithelial cadherin is required to initiate metastatic outgrowth of breast cancer. *Molecular biology of the cell* **22**(14): 2423-2435.

Wilson BG, Roberts CW. 2011. SWI/SNF nucleosome remodellers and cancer. *Nature reviews Cancer* **11**(7): 481-492.

Youmans L, Taylor C, Shin E, Harrell A, Ellis AE, Seguin B, Ji X, Zhao S. 2012. Frequent Alteration of the Tumor Suppressor Gene APC in Sporadic Canine Colorectal Tumors. *PloS one* **7**(12): e50813.

# CHAPTER 3

# CANINE SPONTANEOUS HEAD AND NECK SQUAMOUS CELL CARCINOMAS REPRESENT THEIR HUMAN COUNTERPARTS AT THE MOLECULAR LEVEL[1]

[1]Deli Liu, Huan Xiong, Angela E. Ellis, Nicole C. Northrup, Kevin K. Dobbin, Dong M. Shin, Shaying Zhao. Submitted to *Plos Genetics*, 10/23/2014.

**ABSTRACT**

Spontaneous canine head and neck squamous cell carcinoma (HNSCC) represents an excellent model of human HNSCC but is greatly understudied. To better utilize this valuable resource, we performed its first genome-wide characterization by investigating 12 canine HNSCC cases, of which 9 are oral, via high density array comparative genomic hybridization and RNA-seq. The analyses reveal that these canine cancers faithfully recapitulate many key molecular features of human HNSCC. These include similar genomic copy number abnormality landscapes, analogous sequence mutation patterns, and recurrent alteration of known HNSCC genes (e.g., *MYC, CDKN2A)* and pathways (e.g., cell cycle, mitogenic signaling, TGFβ signaling). Amplification or overexpression of protein kinase genes, matrix metalloproteinase genes, and epithelial–mesenchymal transition genes *TWIST1* and *SNAI* are also prominent in these canine tumors. Our study reemphasizes the value of spontaneous canine cancers in HNSCC basic research (exemplified by our pilot driver-passenger discrimination for human 8q) and anticancer drug discovery.

## INTRODUCTION

During the past several decades, great advances have been made in our understanding of the biology of cancer (Cantley and Baselga 2011). For example, owing to the developments in next-generation sequencing and other high throughput technologies, thousands of human head and neck squamous cell carcinoma (HNSCC) cases have been characterized and significantly altered genes and pathways (involved in cell cycle, mitogenic signaling, etc.) have been identified (Agrawal et al. 2011; Leemans et al. 2011; Stransky et al. 2011; Bose et al. 2013; Pickering et al. 2013). However, translating these research findings into clinical success has been frustratingly slow, and drug development remains a lengthy and expensive process (Hait 2010), with costs currently estimated at over US$1 billion to bring a new drug to market (Hait 2010). One significant challenge is the lack of effective predictive models (Hait 2010). Current preclinical research in drug discovery primarily relies on cell culture and xenograft or genetically-induced rodent models. While very valuable, these models usually do not represent the vast heterogeneity and complexity of human cancers such as HNSCC. Consequently, drug effects in these preclinical cancer models often do not predict clinical results. Thus, a cancer model that can bridge preclinical models and human clinical trials is critically missing and urgently needed.

Spontaneously occurring cancers in pet dogs have numerous advantages (Meuten 2002; Candille et al. 2007; Paoloni and Khanna 2008; Gordon et al. 2009; Tang et al. 2010; Boyko 2011; Rowell et al. 2011; Angstadt et al. 2012; Youmans et al. 2012; Karyadi et al. 2013; Rotroff et al. 2013; Davis and Ostrander 2014; Fenger et al. 2014; Gorden et al. 2014; Liu et al. 2014) that may enable them to bridge this gap. First, unlike genetically-modified or xenograft rodent

models, these cancers are naturally occurring and heterogeneous, capturing the essence of human cancers. Second, as companion animals, dogs share the human environment and are exposed to many of the same carcinogens. For example, canine tonsillar tumors are more common in large cities but rare in rural areas (Reif and Cohen 1971), indicating that environmental chemicals may play a role in tumorigenesis. Third, dogs better resemble humans in biology, e.g., with similar telomere and telomerase activities (Nasir et al. 2001) and frequent occurrence of spontaneous epithelial cancers (Meuten 2002), than mice (Rangarajan and Weinberg 2003). Critically, numerous anatomic and clinical similarities are noted for the same types/subtypes of cancer between the two species and similar treatment schemes are used (Paoloni and Khanna 2008; Gordon et al. 2009; Rowell et al. 2011). Lastly, the large population of pet dogs (~70 million estimated in the US) and the availability of a genome assembly nearly as accurate as the mouse or rat genome (Lindblad-Toh et al. 2005; Ji and Zhao 2008), in contrast to that available for another companion animal, the cat, make canine cancers even more valuable in basic and clinical research.

Spontaneous HNSCCs are relatively frequent in the dog, and the oral cavity is a common site (oral SCC, abbreviated to oSCC hereafter) (Strafuss et al. 1976; Gardner 1996; Meuten 2002). Importantly, many clinical similarities between human and canine HNSCCs have been documented (e.g., metastasis of oSCC to regional lymph nodes) and similar treatments are used (de Vos et al. 2005). However, unlike human HNSCC which is one of the better studied cancers (Klass and Shin 2007; Vermorken et al. 2008; Leemans et al. 2011; Bose et al. 2013; Shin and Khuri 2013), canine HNSCCs remain largely uncharacterized at the molecular level. Our current literature search indicates that not a single canine HNSCC genome or transcriptome has ever

been investigated by sequencing, microarray, or other strategies. Thus, there are insufficient data to evaluate the molecular similarity between these canine cancers and their human counterparts, a key factor to consider in assessing the usage of spontaneous canine cancers in anticancer drug development.

HNSCC represents the sixth leading cancer by incidence in humans, with a half million new cases yearly worldwide (Kamangar et al. 2006); thus, developing effective therapeutic interventions for this disease represents a substantial medical need. To test the hypothesis that spontaneous canine HNSCC can effectively bridge the gap between preclinical models and human clinical trials in anti-HNSCC drug development, we set out to conduct the first genome-wide characterization of spontaneous canine HNSCCs to evaluate the dog-human molecular homology.

## RESULTS

### Twelve spontaneous canine HNSCC, including nine oSCC, cases were investigated.

We characterized 12 spontaneous canine HNSCC cases by genome-wide analyses including high density aCGH and RNA-seq. These cancers come from different dog breeds, including four Labrador retrievers, three mixed breeds, and others. Among the 12 HNSCC cases, nine are oral (oSCCs), with one from the buccal mucosa, three from the tongue, and five from the gingiva. The remaining three cases are from the nasal planum, the nostril, and the eye (the ocular adnexa). The tumors are all invasive (Figure 1A and 1B), and most are well-differentiated

although some are more disorganized than others. Furthermore, two cases appear to be papillomavirus-positive, based on the analysis described in later sections.

**aCGH analysis reveals a strong dog-human homology in genomic copy number abnormality.**

Analogous to human HNSCCs (Leemans et al. 2011; Pickering et al. 2013), aCGH analysis reveals that seven canine tumors harbor extensive copy number abnormalities (CNAs) while the remaining five tumors have hardly any CNAs in their genomes (Figure 2A). In some cases, the variation in the CNA prevalence is clearly related to the cancer progression stage (see tumors 419 and 419-2 in Figure 2A). In other cases however, additional factors may also have contributed. For example, tumor 1152 lacks CNAs but is at a tumor-progression stage similar to those with extensive CNAs (Figure 2A). Interestingly, among the nine oSCCs, those with a buccal mucosa or tongue location harbor significantly more CNAs than those located in the gingiva (Figure 2A).

As in human HNSCCs (Leemans et al. 2011; Pickering et al. 2013), both focal and broad CNAs were detected in the canine tumors (Figure 2A). For example, tumor 240 harbors a focal amplification of ~10Mb located on chromosome 16, increasing the copy number of nearly all 63 genes inside by approximately 2-fold (Figure 2A). Of these genes, 40 are also overexpressed (see later sections), among which "negative regulation of apoptosis" and "endopeptidase activity" are the most enriched functional groups. Indeed, anti-apoptosis associated genes *IKBKB*, *POLB*, *SFRP1* and *FNTA*, as well as endopeptidase genes *ADAM9* and *PLAT* encoded in the region are both amplified and overexpressed. These observations support that this focal amplification event

may have contributed to the pathogenesis of tumor 240. Meanwhile, broad events such as recurrent amplification of canine chromosome 13 were also observed (Figure 2A). The first 40Mb of canine chromosome 13 is syntenic to the last 48Mb of human chromosome 8q, which encodes genes including *MYC* and is one of the most recurrently amplified sites in human HNSCC (Pickering et al. 2013). These observations support common drivers of HNSCC between the human and the dog.

In most canine HNSCCs, more amplifications than deletions were found, causing more genomic sequences and genes to be amplified than deleted (Figure 2A). Amplifications are, on average, also larger than deletions. Importantly, a better correlation between the copy number status and the expression level (see later sections) was observed for amplified genes than deleted genes (Figure 2B), supporting that amplified regions harbor more cancer drivers than deleted regions. This is also consistent with human cancer findings (Pickering et al. 2013).

Many known human HNSCC genes (Chen and Chen 2008; Freier et al. 2010; Pickering et al. 2013; Wei et al. 2013) are also amplified/deleted in the canine tumors. Examples include recurrent amplification of oncogenes *MYC*, *MET*, *PTK2* and *NDRG1*; cell cycle genes *CCNE1*, *CDK6* and *E2F3*. Notably, the most enriched functions among the amplified genes are protein kinase activity, with 15 serine/threonine kinases and 13 tyrosine protein kinases, and protease activity, with 10 serine proteases. This is consistent with TCGA's human HNSCC study. Examples of deleted genes in the canine tumors (Figure 2A) include cell cycle gene *CDKN2A*, one of the best known gene deletions in human HNSCC (Leemans et al. 2011; Pickering et al. 2013), and protein phosphatase *PTPRD*, a tumor suppressor (Ortiz et al. 2014). Furthermore, the

most significantly enriched functions among the deleted genes are closely related to cell polarity including adhesion (18 genes), small GTPase regulator activity (19 genes), and cell junctions (11 genes). This agrees with findings in human cancer and is consistent with the concept that loss of cell polarity is a hallmark of epithelial cancers such as HNSCC (Royer and Lu 2011). These observations support the dog-human molecular homology.

**RNA-seq analysis reveals a strong dog-human homology in transcriptomic alterations.**

To better understand alterations at the transcriptomic level, we performed RNA-seq on seven of the canine oSCCs and three matching normal tissue samples. The study further supports a strong dog-human molecular homology. First, principle component analysis (PCA) separates the tumors from the normal samples (Figure 3A). More importantly, tumors 1172, 465, and 404 are distant from the other tumors in the PCA (Figure 3A), the significance of which will be discussed in later sections. Second, compared to the normal samples, the genes upregulated in the tumors are enriched in functions including: 1) cell adhesion/motility, extracellular matrix, and endopeptidase activity; 2) hypoxia and polysaccharide metabolic processes; 3) blood vessel morphogenesis and cell differentiation; and 4) immune response (Figure 3B). As in human cancers, these functions facilitate canine tumor cell proliferation and invasion.

We examined genes differentially expressed between tumors with extensive CNAs and tumors without (Figure 2A). About 64 genes are upregulated in CNA-rich tumors and are significantly enriched in functions associated with defense response and immune response. About 38 genes are upregulated in CNA-free tumors and are significantly enriched in functions

related to cell morphogenesis and adhesion. These observations agree with our analysis with TCGA human HNSCC data, further supporting the dog-human homology.

We followed published strategies (Cancer Genome Atlas 2012; Pickering et al. 2013) to identify over/under-expressed genes. Consistent with CNA findings (Figure 2A), more genes are overexpressed than underexpressed in all tumors except tumor 404. Importantly, genes recurrently overexpressed among the tumors are significantly enriched in functions associated with cell cycle (e.g., cytoskeleton, spindle, centrosome, kinetochore, etc.), protein kinase activity (e.g., *PTK2*, *TEC*, *CHEK2*, etc.), nucleolus, and mRNA and ncRNA processing. Recurrently underexpressed genes are, however, significantly enriched in functions related to cell junctions (e.g., 9 tight junction genes), mitochondria (e.g., respiratory chain and oxidative phosphorylation), serine protease inhibitors, and apoptosis. These functions promote cancer cell proliferation and invasion, consistent with human cancer findings.

Critically, we found the same genes and pathways altered in these canine cancers as reported in human oSCCs (Pickering et al. 2013). For example, gene members of the mitogenic pathway such as *EGFR*, *PIK3CA*, *BRAF*, and *AKT1* are recurrently overexpressed among the canine tumors (Figure 3C). *AKT1* is especially noteworthy because its expression level in each tumor is consistently higher than in each normal sample by 2-4-fold (Figure 3C). Cell cycle is also altered, as evidenced by overexpression of multiple cyclin genes, *CDK4*, *CDK6* and *E2F1;* as well as underexpression of *CDKN2A* and *CDKN2B* (Figure 3C). These pathway alterations could promote cancer cell proliferation.

Another pathway affected is TGFβ signaling, as evidenced by the recurrent overexpression of *TGFB1*, *TGFB2*, *TGFBR2*, *TGFBR1*, *SMAD3* and *SMAD4* in the canine tumors (Figure 3C). Other notable changes include at least 12 matrix metalloproteinase (MMP) genes, whose expression increased by hundreds to thousands fold in at least one tumor (Figure 3C). Likewise, epithelial to mesenchymal transition (EMT) genes *TWIST1* and *SNAI1* are also recurrently overexpressed (Figure 3C). Activation of these genes would facilitate the invasion of tumor cells into adjacent tissues. Interestingly, tumor 1172 appears to be an exception to this gene overexpression profile (which will be revisited in the DISCUSSION section).

Finally, two canine oSCCs appear to be canine papillomavirus (CPV)-positive, based on the detection of papillomavirus sequences, specifically CPV7 E2/E4 sequences in tumor 1172 and sequences with high homology (>90%) to human papillomavirus HPV77 E2/E4 in tumor 465, among their RNA-seq reads.

**RNA-seq analysis reveals dog-human homologies in sequence mutation.**

We took advantage of RNA-seq data to examine sequence mutations in the canine samples. Briefly, to achieve more accurate mutation-finding, we utilized only coding regions with 30-300X RNA-seq read coverage, which distribute across the genome and amount to 4-6 Mb sequences in total per sample. The analysis again reveals dog-human homologies. First, base transitions C↔T/G↔A dominate base transversions in all samples (Figure 4A), indicating similar mutation mechanisms in both species. The base substitution rate ranges from 214 to 273 per Mb coding sequences in these samples, with tumor 404 having the highest rate (which will be revisited in DISCUSSION). Second, the analysis uncovered a somatic mutation, E233K, in

*TP53*. Similarly, consistent with human studies (Agrawal et al. 2011; Stransky et al. 2011), genes *FAT1*, *FAT2*, *UBR2*, *TNC* and others were found to be mutated in the canine tumors.

As an example, Figure 4B shows the mutations found in the gene *RELN* in canine tumor 404, including one nonsense, five non-synonymous and one synonymous changes. Importantly, *RELN* is significantly mutated in human HNSCC (Agrawal et al. 2011), non-small cell lung cancer of smokers (Govindan et al. 2012), and acute lymphoblastic leukemia (Zhang et al. 2012). *RELN* encodes an extracellular matrix glycoprotein which is known to control cell–cell interactions to regulate neuronal migration and positioning in the developing brain. Thus, alteration of *RELN* may contribute to tumor cell invasion and spread in both the human and the dog.

**Driver-passenger discrimination via dog-human comparison for human 8q.**

The strong dog-human homologies described above provide the fundamental justification for applying the dog-human comparison strategy for driver-passenger discrimination as described (Tang et al. 2014) for HNSCC. Because of the small sample size of canine tumors, we tested this strategy only on human 8q, one of the most recurrently amplified regions in human HNSCC (Pickering et al. 2013). Due to interspecies genomic rearrangements, human 8q is broken into two dog chromosomal regions, which include chromosome 29 and the first 38Mb of chromosome 13 (Figure 5A). Notably, the entire human 8q is significantly amplified among TCGA's 948 human HNSCCs (FDR < $10^{-6}$), leading to the amplification of all of its 398 genes (Figure 5A). In canine tumors, however, only the chromosome 13 region is significantly amplified, resulting in the amplification of 125 genes out of 210 total at FDR < 0.2. For

chromosome 29 in contrast, merely 2 genes out of 188 total are amplified. These numbers significantly differ ($p < 2.2 \times 10^{-16}$) (Figure 5A). Thus, based on our strategy (Tang et al. 2014), amplified genes (125 total) on chromosome 13 are considered as driver candidate genes (DCGs), while unchanged genes on chromosome 29 (186 total) are deemed passenger candidate genes (PCGs) (Figure 5A).

We then performed several analyses to examine the differences between the DCGs and PCGs identified. First, a significantly better correlation between the copy number status and the mRNA expression level was observed for DCGs than PCGs, using data from TCGA's HNSCC project and the oesophagus study from the cancer cell line encyclopedia (CCLE) (Barretina et al. 2012) (Figure 5B). This indicates that amplification of DCGs is more functionally relevant than amplification of PCGs. Second, significantly more DCGs are mutated than PCGs, based on published human HNSCC studies (Stransky et al. 2011) ($p < 0.0081$). Lastly, well known cancer driver genes such as *MYC* are among the DCGs.

## DISCUSSION

In this study, we performed the first genome-wide characterization of spontaneous canine HNSCC. Although the sample size (12 cases including 9 oral) is small, the study reveals a strong dog-human molecular homology at various levels and reinforces the value of spontaneous canine cancers in HNSCC basic and clinical research.

**Spontaneous canine HNSCCs faithfully represent their human counterparts at the molecular level.**

As described throughout the RESULTS section, our genomic and transcriptomic studies reveal strong dog-human molecular homologies for HNSCC at various levels. For large genomic changes, the two species share a similar CNA landscape, including large amplicons that likely harbor cancer drivers. At the individual gene level, many of the same amplified/deleted or over/under-expressed genes are found between the two species, and the altered genes are enriched in the same functional groups. Examples include protein kinase or protease activity for amplified/overexpressed genes, as well as cell adhesion and other epithelial polarity related functions for deleted/underexpressed genes. The high level overexpression of EMT genes and MMP genes are especially noteworthy. At the pathway level, both species show alterations in cell cycle, mitogenic signaling, and TGFβ signaling. These gene and pathway alterations contribute to cancer development and progression in both the dog and the human.

The dominance of C↔T changes over other base substitution types indicates that deamination of C to U/T is a major sequence mutation mechanism in dogs as in humans. This result is consistent with aging being a risk factor for HNSCC development in both species (Alexandrov et al. 2013). Canine tumor 404 is especially noteworthy in this regard, as it lacks genomic CNAs but has the greatest number of underexpressed genes and the highest sequence mutation rate. Hence, tumor 404 likely belongs to the subtype that displays hyper-mutation and hyper-methylation in CpG islands, with 5mC→T changes as the mutation mechanism, as reported in human cancers (Cancer Genome Atlas 2012).

Our study indicates that canine tumors 1172 and 465 are CPV-positive. Analogous to human HNSCCs (Leemans et al. 2011), both tumors (especially tumor 1172) are distinct from other canine oSCCs in gene expression. Importantly, consistent with the finding that HPVs infect the basal layer of squamous epithelium (Blitzer et al. 2014), tumor 1172 exhibits features indicating basal stem cell origin. Unlike other canine oSCCs, tumor 1172 overexpresses the pluripotent marker *SOX2* and at least 20 homeobox genes that are associated with embryonic morphogenesis, but does not overexpress EMT genes or many of the MMP genes.

We hypothesize the carcinogenic mechanism of the canine tumors based on our findings (Figure 6). Regarding cells of origin, unlike tumor 1172 discussed above, we propose that other tumors arise from more differentiated cells of the squamous epithelium (Figure 6), because of overexpression of protease genes and EMT genes. At the genome level, we hypothesize that primary drivers include focal amplifications for tumor 240, but DNA methylation and sequence mutation for tumor 404 (Figure 6). Finally, notable gene alterations and recurrently altered pathways are listed as cancer drivers, as shown in Figure 6.

**Spontaneous canine HNSCCs are valuable for efficient cancer driver-passenger discrimination.**

The strong dog-human molecular homology described above fundamentally justifies the use of our established dog-human comparison strategy for driver-passenger discrimination (Tang et al. 2014) for HNSCC. Indeed, our pilot human 8q study has shown that this approach is valid. The use of canine cancers in this regard is highly significant. For example, our analysis with the copy number data of TCGA's 948 human HNSCC cases has found 3,493 amplified genes and

3,852 deleted genes at FDR ≤ 0.05, including those harbored by both focal amplifications/deletions and broad (chromosomal arm level) gains/losses. Among these genes, some are drivers and some are passengers. Based on our analyses, a sample size of 948 tumors is already saturating, and studying additional human tumors no longer helps in determining which amplified/deleted genes are drivers and which are passengers. However, investigating about 90 canine tumors will reduce the number of driver candidates by at least half, according to our estimation, significantly reducing the workload of downstream functional validation which is time-consuming and expensive.

**Spontaneous canine HNSCCs can bridge the gap between preclinical models and human clinical trials, accelerating anti-HNSCC drug development.**

Although our understanding of the molecular and cellular biology of cancer including HNSCC has been greatly advanced in the past several decades, translating these research findings into clinical success has been frustratingly slow. One major obstacle is the lack of effective predictive models (Hait 2010). Current preclinical models, including cell culture and xenograft or genetically-induced rodent models, typically fail to represent the vast heterogeneity and complexity of human cancers and often do not predict clinical results. Thus, a cancer model that can bridge the gap between these preclinical models and human clinical trials is critically missing and urgently needed.

Spontaneous canine HNSCCs can serve as such a translational model, as our study indicates that they faithfully represent their human counterparts at the molecular level. By effectively bridging the gap between the preclinical models and human clinical trials, coupled with less stringent FDA regulations governing pet clinical trials, these canine cancers can

significantly accelerate the translation of basic research findings into successful clinical applications. For example, for drugs targeting *AKT1* and other genes in the mitogenic pathway (Bendell et al. 2012; Gordon and Banerji 2013; Klempner et al. 2013; Shin and Khuri 2013), their efficacy, toxicity, dosage, and schedule can be more accurately evaluated through clinical trials with canine patients before entering human clinical trials. This will significantly reduce the cost and accelerate new anticancer drug discovery. Likewise, many other druggable genes such as protein kinases (e.g., *LYN*, *SRC*, *PTK2, etc.*) are recurrently activated in canine oSCC (Figure 7), analogous to human oSCC (Pickering et al. 2013). New drugs targeting these kinase activations can be first investigated in canine cancer patients.

Finally, we hope that our study will significantly raise the awareness of spontaneous cancers in pet dogs, an enormously valuable but greatly understudied and underutilized resource.

## MATERIALS AND METHODS

### Canine Tissue Samples

Fresh-frozen normal and tumor tissue samples of spontaneous canine HNSCCs were obtained from the Animal Cancer Tissue Repository of Colorado State University. Samples were collected from client-owned dogs that developed the disease spontaneously, under the guidelines of the Institutional Animal Care and Use Committee and with owner informed consent.

**Tissue Dissection, DNA and RNA Extraction, and PCR Analyses**

Tissue cryosectioning, H&E staining, and cryomicrodissection were performed as described (Tang et al. 2010) to enrich for tumor cells in tumor samples and squamous epithelial cells in normal samples. Genomic DNA and RNA were then extracted from the dissected tissues using the AllPrep DNA/RNA Mini Kit (cat. no. 80204) from Qiagen. Only samples with a 260/280 ratio of ~1.8 (DNA) or ~2.0 (RNA) and showing neither degradation nor other contamination on the agarose gels were subjected to further analyses.

**aCGH Analyses**

Canine aCGH experiments were conducted at the Florida State University Microarray Facility, with 385K canine CGH array chips from Roche NimbleGen Systems, Inc. CNAs were identified as described (Tang et al. 2010). Significantly amplified/deleted genes were identified by GISTIC (Beroukhim et al. 2007) for both our canine tumors and the 948 human HNSCCs downloaded from TCGA site (cancergenome.nih.gov/).

**Paired-end RNA-seq**

Sequencing was conducted using the Illumina platform, following the protocols from the manufacturer. RNA-seq was performed at the HudsonAlpha Institute for Biotechnology or the BGI-America, yielding 48 to 66 million paired-end sequence reads of 50bp or 49bp per sample.

**RNA-seq Data Analysis**

RNA-seq data analyses were performed as described (Liu et al. 2014). Briefly, read pairs were aligned to the dog reference genome (Lindblad-Toh et al. 2005) with TopHat v2.0.5

(tophat.cbcb.umd.edu). The uniquely mapped pairs were used to quantify a gene's expression level by calculating its FPKM (fragments per kilobase of exon per million mapped fragments) value, using Cufflinks (cufflinks.cbcb.umd.edu) with default parameters and the canine gene annotation downloaded from the University of Santa Cruz (UCSC) genome site. Base substitutions were identified with VarScan2 (varscan.sourceforge.net) in coding regions with RNA-seq read coverage ranging from 30X to 300X. Over/underexpressed genes in cancers were identified as described (Cancer Genome Atlas 2012; Pickering et al. 2013). Differentially expressed genes between two groups of samples were identified by DESeq (Anders and Huber 2010) and t-tests. Gene functional annotation and enrichment analyses were achieved using the DAVID Functional Annotation tool at david.abcc.ncifcrf.gov.

**Data Access**

RNA-seq data have been submitted to the NCBI SRA database under accession number SRP046723. aCGH data have been submitted to the GEO database under accession number GSE61231.

**Acknowledgements**

We thank Mr**.** Roger Nilsen and the Georgia Genomics Facility for making the RNA-seq libraries; Dr. Shawn Levy at the HudsonAlpha Institute for Biotechnology and the BGI for sequencing; and Dr. Steven H. Miller at the FSU-NimbleGen Microarray Facility for the aCGH work; Ms. Teri Guerrero, Ms. Irene Mok and Dr. Susan E Lana for their help on dog sample collection; and Drs. Georgia Chen and Mohammad Aminur Rahman for insightful discussions on the study.

## FIGURE LEGENDS

**Figure 1. Representative H&E stained images of canine normal squamous epithelium and SCC of the oral cavity.**

**A**, normal squamous epithelium of the oral mucosa (from case 240). Epithelium is between the arrow and the arrowhead. The arrow indicates the basal layer, while the arrowhead designates the stratum corneum.

**B**, SCC of the oral mucosa (from case 240). The arrow points to an area with loss of a distinct basal layer and basement membrane. The arrowhead indicates keratinizing squamous cells. The images are at 100X magnification.

**Figure 2. aCGH analysis reveals extensive CNAs in the majority of canine HNSCCs**.

**A**, CNAs found in canine HNSCCs. The images were drawn as described (Tang et al. 2010), with each line representing a canine chromosome and vertical lines above/below the chromosome indicating amplifications (red) or deletions (blue) respectively. Notable amplified (red) or deleted (blue) genes, as well as tumor IDs (e.g., 240) and locations (e.g., buccal mucosa) are indicated. Tumors 419 and 419-2 are from the same case (ID 419), with 419-2 being less advanced. One focal amplification (20.8-30.7Mb of chromosome 16) in tumor 240 is shown to exemplify that many of the amplified genes are also overexpressed and associated with tumor-promoting function.

**B**, integration of genes' copy number status and mRNA expression level. Genes (11,821 total) with FPKM (fragments per kilobase of exon per million mapped fragments) of $\geq 1$ in at least one of the tumors subjected to RNA-seq were sorted into three groups based on the analysis indicated in (**A)**: deleted (Del), unchanged, or amplified (Amp) as shown on the X-axis. Then, each group

of genes was further divided into three subgroups based on the mRNA expression level: low, normal, and high determined as described (Pickering et al. 2013), with the percentages shown in the Y-axis.

**Figure 3. The same pathways are altered in canine oSCCs as in their human counterparts as revealed by RNA-seq analysis.**

**A**, PCA with FPKM values of 12,478 total genes, each having FPKM ≥ 1 in at least one of the tumor (with ID followed by T, e.g., 240T) or normal (with ID followed by "N", e.g., 240N) samples. The plot shows that tumors are separated from normal samples and tumors 1172T, 404T, and 465T are distinct from the other tumors.

**B**, heatmap with the $\log_2(FPKM)$ values of genes (255 total) that are differentially expressed between the tumors and the normal samples found by both DESeq and t-tests at FDR ≤ 0.1. The right panel illustrates the significantly enriched functions of each gene cluster indicated. Red denotes upregulation and green denotes downregulation.

**C**, heatmap of genes from pathways and groups that are reported to be altered in human oSCC (Pickering et al. 2013). The heatmap is generated as described for (**B)**.

**Figure 4. RNA-seq analysis reveals dog-human homology in sequence mutations.**

**A**, the fractions (the Y-axis) of base substitution types in the canine samples (the X-axis) detected by RNA-seq when compared to the dog reference genome (Lindblad-Toh et al. 2005).

**B**, synonymous (green dots) and non-synonymous (yellow dots) substitutions, and a nonsense mutation (red star) uncovered in the *RELN* gene in tumor 404.

**Figure 5. Pilot HNSCC driver–passenger discrimination via human–dog comparison for human 8q.**

**A**, driver-passenger discrimination of human 8q. Human 8q is broken into dog chromosomes 29 (chr29) and 13 (chr13), with the numbers indicating the sequence coordinates. In human HNSCCs, the entire human 8q and all genes encoded (398 total) are recurrently amplified, as represented by a broken lined bar. In canine tumors, however, only chr13 is significantly amplified resulting in 125 genes being amplified and 85 genes unchanged, compared to only 2 genes amplified and 186 genes unchanged for chr29. The 125 amplified genes on chr13, including *MYC*, are considered as driver candidate genes (DCGs; in green), whereas the 186 unchanged genes on chr29 are deemed passenger candidate genes (PCGs; in red).

**B**, the correlation between copy number status, represented by a gene's $\log_2 \frac{Tumor}{Normal}$ value, and mRNA expression level, represented by a gene's $\log_2(FPKM)$ from RNA-seq or $\log_2(intensity)$ from microarray. Two data sources, TCGA and CCLE, were used as indicated.

**Figure 6**. **The proposed carcinogenic mechanisms of canine oSCCs investigated**. The cells of origin in these tumors are hypothesized as indicated, with oval cells representing the basal stem cells. Alterations at various levels are shown. For individual gene alterations, red represents gene amplification or overexpression, while green represents gene deletion or underexpression.

**Figure 7**. **Many druggable genes are overexpressed in canine oSCC**. The heatmap was generated as described for Figure 3C. The gene list and the therapeutic agents targeting each gene alteration are obtained from a published human oSCC study.

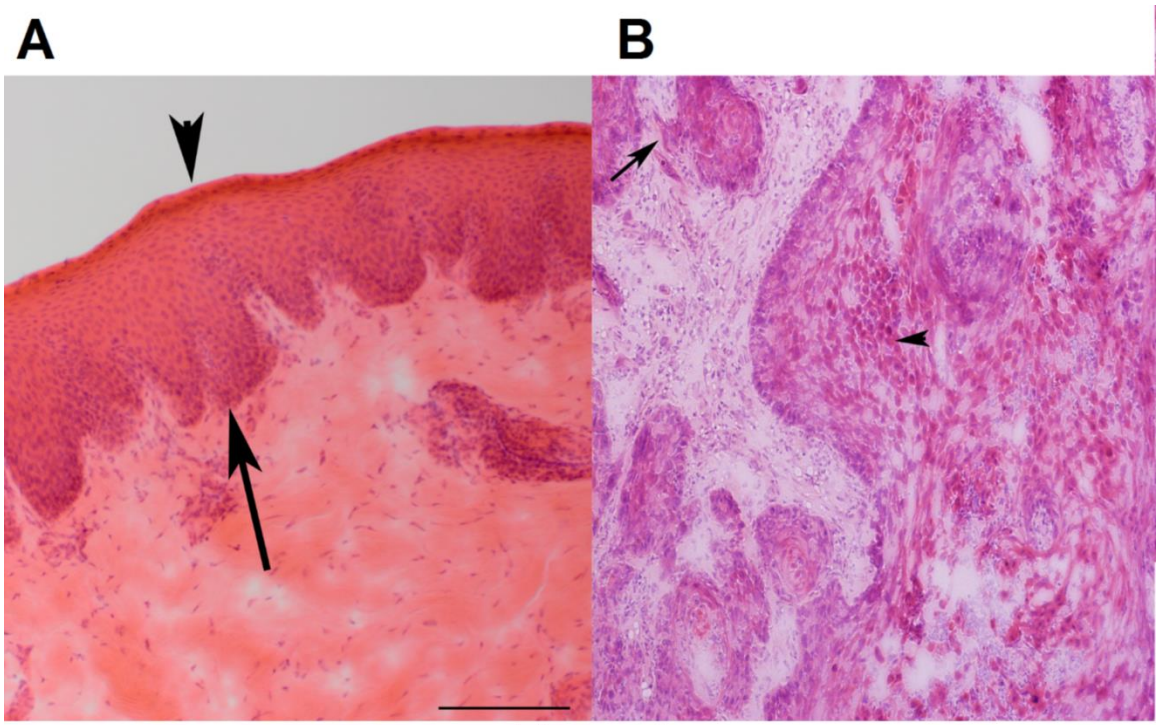



**Figure 1**

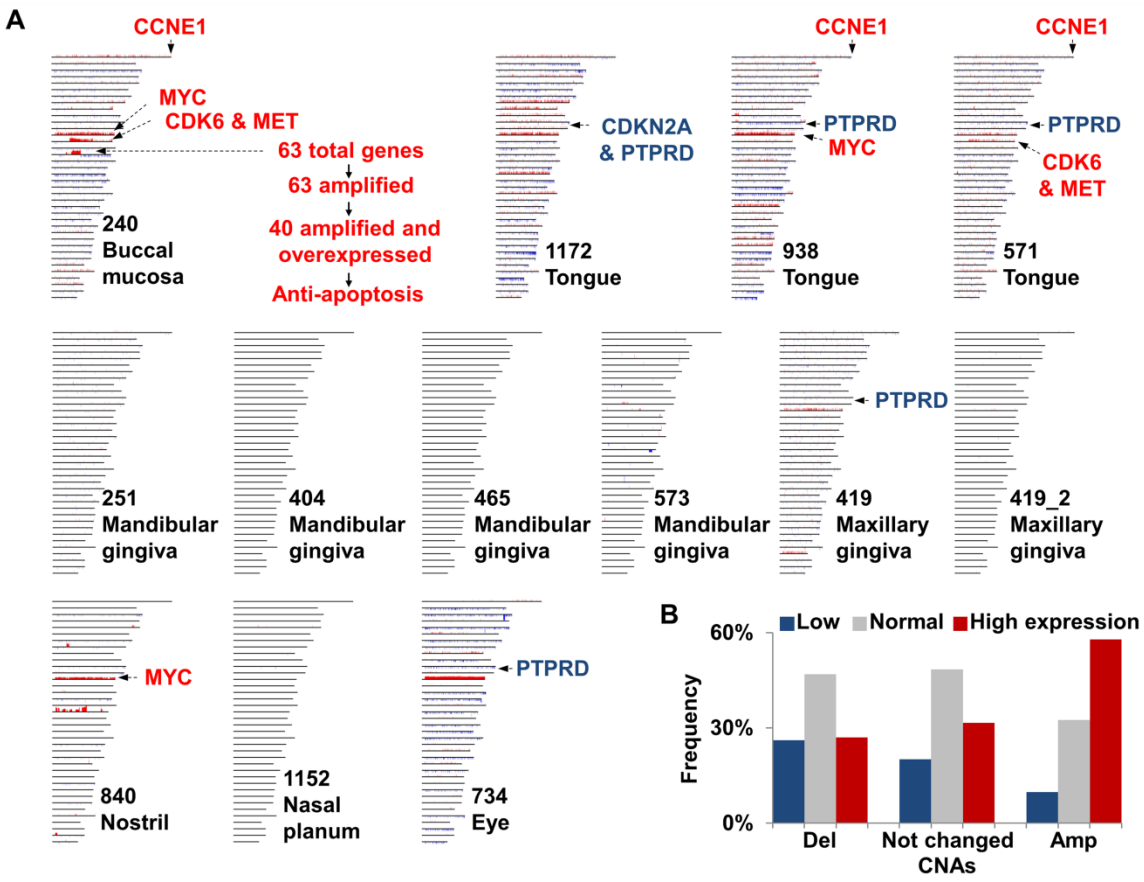

A
CCNE1
MYC
CDK6 & MET
63 total genes
63 amplified
40 amplified and overexpressed
Anti-apoptosis
240
Buccal mucosa
CDKN2A & PTPRD
1172
Tongue
CCNE1
PTPRD
MYC
938
Tongue
CCNE1
PTPRD
CDK6 & MET
571
Tongue
251
Mandibular gingiva
404
Mandibular gingiva
465
Mandibular gingiva
573
Mandibular gingiva
PTPRD
419
Maxillary gingiva
419_2
Maxillary gingiva
MYC
840
Nostril
1152
Nasal planum
PTPRD
734
Eye
B
Low
Normal
High expression
60%
30%
0%
Frequency
Del
Not changed CNAs
Amp

**Figure 2**

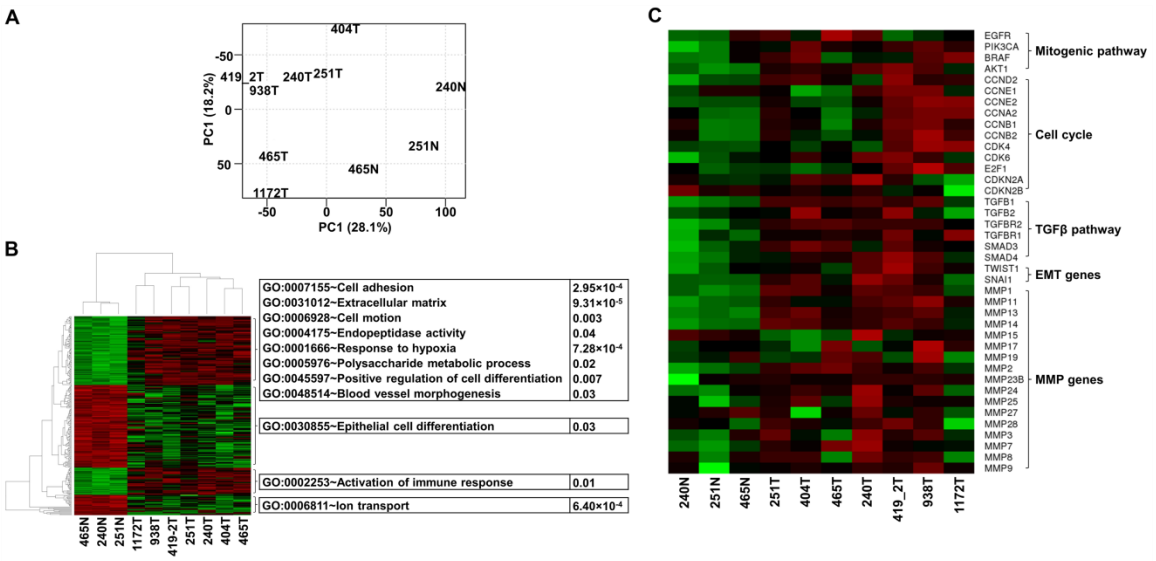

A
404T
-50
419_2T
240T
251T
938T
240N
0
251N
50
465T
465N
1172T
-50
0
50
100
PC1 (28.1%)
PC1 (18.2%)
B
GO:0007155~Cell adhesion 2.95×10-4
GO:0031012~Extracellular matrix 9.31×10-5
GO:0006928~Cell motion 0.003
GO:0004175~Endopeptidase activity 0.04
GO:0001666~Response to hypoxia 7.28×10-4
GO:0005976~Polysaccharide metabolic process 0.02
GO:0045597~Positive regulation of cell differentiation 0.007
GO:0048514~Blood vessel morphogenesis 0.03
GO:0030855~Epithelial cell differentiation 0.03
GO:0002253~Activation of immune response 0.01
GO:0006811~Ion transport 6.40×10-4
465N
240N
251N
1172T
938T
419-2T
251T
240T
404T
465T
C
EGFR
PIK3CA
BRAF
AKT1
Mitogenic pathway
CCND2
CCNE1
CCNE2
CCNA2
CCNB1
CCNB2
CDK4
CDK6
E2F1
CDKN2A
CDKN2B
Cell cycle
TGFB1
TGFB2
TGFBR2
TGFBR1
SMAD3
SMAD4
TGFβ pathway
TWIST1
SNAI1
EMT genes
MMP1
MMP11
MMP13
MMP14
MMP15
MMP17
MMP19
MMP2
MMP23B
MMP24
MMP25
MMP27
MMP28
MMP3
MMP7
MMP8
MMP9
MMP genes
240N
251N
465N
251T
404T
465T
240T
419_2T
938T
1172T

**Figure 3**

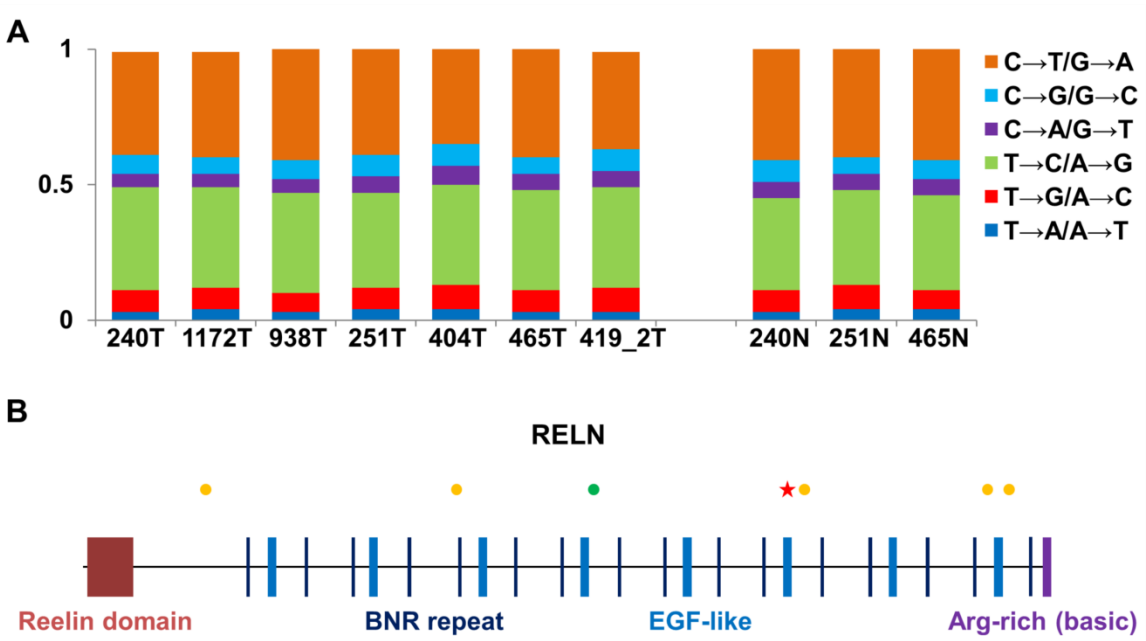
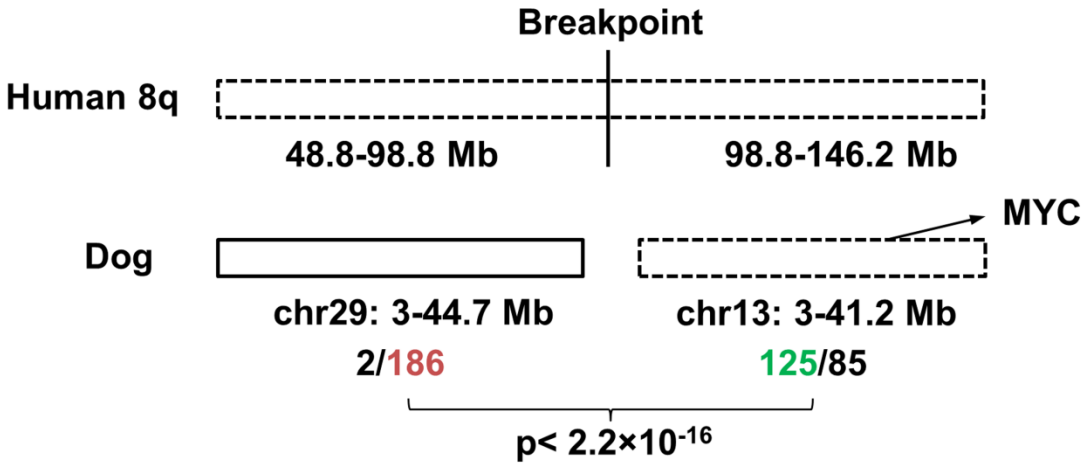

A
1
0.5
0
240T
1172T
938T
251T
404T
465T
419_2T
240N
251N
465N
C→T/G→A
C→G/G→C
C→A/G→T
T→C/A→G
T→G/A→C
T→A/A→T
B
RELN
Reelin domain
BNR repeat
EGF-like
Arg-rich (basic)

Figure 4

A

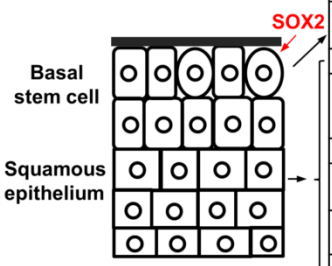



B

Correlation between copy number and expression

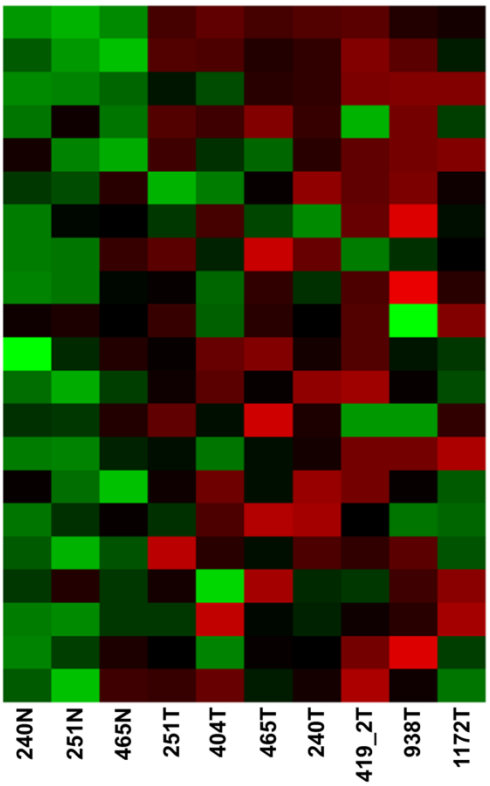
| | PCGs | DCGs | P-value |
|---|---|---|---|
| TCGA HNSCC | 0.0324 | 0.0579 | 0.0001 |
| CCLE oesophagus | 0.0833 | 0.1044 | 0.0389 |

Figure 5




| | Genomic alteration | CPV/HPV | Genes | Signaling pathway |
|---|---|---|---|---|
| 1172T | CNAs | CPV7 | SOX2, CDKN2B, PTPRD | Mitogenic, cell cycle |
| 419_2T | - | - | MMPs, TWIST1, SNAI1 | Mitogenic, cell cycle, TGFβ |
| 938T | CNAs | - | CCNE1, MYC, MMPs, TWIST1, SNAI1, PTPRD | Mitogenic, cell cycle, TGFβ |
| 251T | - | - | MMPs, TWIST1, SNAI1 | Mitogenic, cell cycle, TGFβ |
| 404T | Sequence mutations, CpG methylation? | - | MMPs, TWIST1, SNAI1 | Mitogenic, TGFβ |
| 465T | - | HPV77 | EGFR, MMPs | Mitogenic, TGFβ |
| 240T | Focal amplifications, TP53 mutation | - | CCNE1, MYC, CDK6, MET, MMPs, TWIST1, SNAI1 | Mitogenic, TGFβ |

**Figure 6**




| Gene | Targeted therapeutics |
|---|---|
| LYN | Dasatinib, bosutinib, sunitinib, vandetanib |
| SRC | Dasatinib, bosutinib, vandetanib |
| STK3 | Sunitinib |
| MAP3K9 | Pazopanib |
| MELK | Sunitinib |
| ABL1 | Dasatinib, imatinib, nilotinib, bosutinib, vandetanib |
| ABL2 | Dasatinib, nilotinib, bosutinib |
| EGFR | Gefitinib, vandetanib, cetuximab |
| EPHB4 | Dasatinib |
| ERBB2 | Lapatinib, trastuzumab, pertuzumab |
| FLT4 | Pazopanib, sorafenib, sunitinib |
| NFKB1 | Bortezomib |
| NTRK2 | Sunitinib |
| PTK2 | Sunitinib |
| STK17A | Sunitinib |
| TNK2 | Dasatinib, bosutinib |
| VEGFA | Bevacizumab |
| CLK2 | Sunitinib |
| HSP90AA1 | Vorinostat, romidepsin |
| MTOR | Everolimus |
| ZAK | Sorafenib |

**Figure 7**

## REFERENCES


Agrawal N, Frederick MJ, Pickering CR, Bettegowda C, Chang K, Li RJ, Fakhry C, Xie TX, Zhang J, Wang J et al. 2011. Exome sequencing of head and neck squamous cell carcinoma reveals inactivating mutations in NOTCH1. *Science* **333**(6046): 1154-1157.

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL et al. 2013. Signatures of mutational processes in human cancer. *Nature* **500**(7463): 415-421.

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome biology* **11**(10): R106.

Angstadt AY, Thayanithy V, Subramanian S, Modiano JF, Breen M. 2012. A genome-wide approach to comparative oncology: high-resolution oligonucleotide aCGH of canine and human osteosarcoma pinpoints shared microaberrations. *Cancer genetics* **205**(11): 572-587.

Barretina J, Caponigro G, Stransky N, Venkatesan K, Margolin AA, Kim S, Wilson CJ, Lehar J, Kryukov GV, Sonkin D et al. 2012. The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**(7391): 603-607.

Bendell JC, Rodon J, Burris HA, de Jonge M, Verweij J, Birle D, Demanse D, De Buck SS, Ru QC, Peters M et al. 2012. Phase I, dose-escalation study of BKM120, an oral pan-Class I PI3K inhibitor, in patients with advanced solid tumors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **30**(3): 282-290.

Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, Vivanco I, Lee JC, Huang JH, Alexander S et al. 2007. Assessing the significance of chromosomal aberrations in

cancer: methodology and application to glioma. *Proceedings of the National Academy of Sciences of the United States of America* **104**(50): 20007-20012.

Blitzer GC, Smith MA, Harris SL, Kimple RJ. 2014. Review of the clinical and biologic aspects of human papillomavirus-positive squamous cell carcinomas of the head and neck. *International journal of radiation oncology, biology, physics* **88**(4): 761-770.

Bose P, Brockton NT, Dort JC. 2013. Head and neck cancer: from anatomy to biology. *International journal of cancer Journal international du cancer* **133**(9): 2013-2023.

Boyko AR. 2011. The domestic dog: man's best friend in the genomic era. *Genome biology* **12**(2): 216.

Cancer Genome Atlas N. 2012. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**(7407): 330-337.

Candille SI, Kaelin CB, Cattanach BM, Yu B, Thompson DA, Nix MA, Kerns JA, Schmutz SM, Millhauser GL, Barsh GS. 2007. A -defensin mutation causes black coat color in domestic dogs. *Science* **318**(5855): 1418-1423.

Cantley LC, Baselga J. 2011. The Era of Cancer Discovery. *Cancer discovery* **1**(1): 1-U15.

Chen Y, Chen C. 2008. DNA copy number variation and loss of heterozygosity in relation to recurrence of and survival from head and neck squamous cell carcinoma: a review. *Head & neck* **30**(10): 1361-1383.

Davis BW, Ostrander EA. 2014. Domestic dogs and cancer research: a breed-based genomics approach. *ILAR journal / National Research Council, Institute of Laboratory Animal Resources* **55**(1): 59-68.

de Vos JP, Burm AG, Focker AP, Boschloo H, Karsijns M, van der Waal I. 2005. Piroxicam and carboplatin as a combination treatment of canine oral non-tonsillar squamous cell

carcinoma: a pilot study and a literature review of a canine model of human head and neck squamous cell carcinoma. *Veterinary and comparative oncology* **3**(1): 16-24.

Fenger JM, London CA, Kisseberth WC. 2014. Canine osteosarcoma: a naturally occurring disease to inform pediatric oncology. *ILAR journal / National Research Council, Institute of Laboratory Animal Resources* **55**(1): 69-85.

Freier K, Knoepfle K, Flechtenmacher C, Pungs S, Devens F, Toedt G, Hofele C, Joos S, Lichter P, Radlwimmer B. 2010. Recurrent copy number gain of transcription factor SOX2 and corresponding high protein expression in oral squamous cell carcinoma. *Genes, chromosomes & cancer* **49**(1): 9-16.

Gardner DG. 1996. Spontaneous squamous cell carcinomas of the oral region in domestic animals: a review and consideration of their relevance to human research. *Oral diseases* **2**(2): 148-154.

Gorden BH, Kim JH, Sarver AL, Frantz AM, Breen M, Lindblad-Toh K, O'Brien TD, Sharkey LC, Modiano JF, Dickerson EB. 2014. Identification of Three Molecular and Functional Subtypes in Canine Hemangiosarcoma through Gene Expression Profiling and Progenitor Cell Characterization. *American Journal of Pathology* **184**(4): 985-995.

Gordon I, Paoloni M, Mazcko C, Khanna C. 2009. The Comparative Oncology Trials Consortium: using spontaneously occurring cancers in dogs to inform the cancer drug development pathway. *PLoS medicine* **6**(10): e1000161.

Gordon V, Banerji S. 2013. Molecular pathways: PI3K pathway targets in triple-negative breast cancers. *Clinical cancer research : an official journal of the American Association for Cancer Research* **19**(14): 3738-3744.

Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, Maher CA, Fulton R, Fulton L, Wallis J et al. 2012. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* **150**(6): 1121-1134.

Hait WN. 2010. Anticancer drug development: the grand challenges. *Nature reviews Drug discovery* **9**(4): 253-254.

Ji X, Zhao S. 2008. DA and Xiao-two giant and composite LTR-retrotransposon-like elements identified in the human genome. *Genomics* **91**(3): 249-258.

Kamangar F, Dores GM, Anderson WF. 2006. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **24**(14): 2137-2150.

Karyadi DM, Karlins E, Decker B, vonHoldt BM, Carpintero-Ramirez G, Parker HG, Wayne RK, Ostrander EA. 2013. A copy number variant at the KITLG locus likely confers risk for canine squamous cell carcinoma of the digit. *PLoS genetics* **9**(3): e1003409.

Klass CM, Shin DM. 2007. Current status and future perspectives of chemoprevention in head and neck cancer. *Current cancer drug targets* **7**(7): 623-632.

Klempner SJ, Myers AP, Cantley LC. 2013. What a Tangled Web We Weave: Emerging Resistance Mechanisms to Inhibition of the Phosphoinositide 3-Kinase Pathway. *Cancer discovery* **3**(12): 1345-1354.

Leemans CR, Braakhuis BJ, Brakenhoff RH. 2011. The molecular biology of head and neck cancer. *Nature reviews Cancer* **11**(1): 9-22.

Lindblad-Toh K Wade CM Mikkelsen TS Karlsson EK Jaffe DB Kamal M Clamp M Chang JL Kulbokas EJ, 3rd Zody MC et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* **438**(7069): 803-819.

Liu D, Xiong H, Ellis AE, Northrup NC, Rodriguez CO, Jr., O'Regan RM, Dalton S, Zhao S. 2014. Molecular homology and difference between spontaneous canine mammary cancer and human breast cancer. *Cancer research*.

Meuten DJ. 2002. *Tumors in domestic animals*. Iowa State University Press, Ames, Iowa.

Nasir L, Devlin P, McKevitt T, Rutteman G, Argyle DJ. 2001. Telomere lengths and telomerase activity in dog tissues: a potential model system to study human telomere and telomerase biology. *Neoplasia* **3**(4): 351-359.

Ortiz B, White JR, Wu WH, Chan TA. 2014. Deletion of Ptprd and Cdkn2a cooperate to accelerate tumorigenesis. *Oncotarget* **5**(16): 6976-6982.

Paoloni M, Khanna C. 2008. Translation of new cancer treatments from pet dogs to humans. *Nature reviews Cancer* **8**(2): 147-156.

Pickering CR, Zhang J, Yoo SY, Bengtsson L, Moorthy S, Neskey DM, Zhao M, Ortega Alves MV, Chang K, Drummond J et al. 2013. Integrative genomic characterization of oral squamous cell carcinoma identifies frequent somatic drivers. *Cancer discovery* **3**(7): 770-781.

Rangarajan A, Weinberg RA. 2003. Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nature reviews Cancer* **3**(12): 952-959.

Reif JS, Cohen D. 1971. The environmental distribution of canine respiratory tract neoplasms. *Archives of environmental health* **22**(1): 136-140.

Rotroff DM, Thomas R, Breen M, Motsinger-Reif AA. 2013. Naturally occuring canine cancers: powerful models for stimulating pharmacogenomic advancement in human medicine. *Pharmacogenomics* **14**(16): 1929-1931.

Rowell JL, McCarthy DO, Alvarez CE. 2011. Dog models of naturally occurring cancer. *Trends in molecular medicine* **17**(7): 380-388.

Royer C, Lu X. 2011. Epithelial cell polarity: a major gatekeeper against cancer? *Cell death and differentiation* **18**(9): 1470-1477.

Shin DM, Khuri FR. 2013. Advances in the management of recurrent or metastatic squamous cell carcinoma of the head and neck. *Head & neck* **35**(3): 443-453.

Strafuss AC, Cook JE, Smith JE. 1976. Squamous cell carcinoma in dogs. *Journal of the American Veterinary Medical Association* **168**(5): 425-427.

Stransky N, Egloff AM, Tward AD, Kostic AD, Cibulskis K, Sivachenko A, Kryukov GV, Lawrence MS, Sougnez C, McKenna A et al. 2011. The Mutational Landscape of Head and Neck Squamous Cell Carcinoma. *Science* **333**(6046): 1157-1160.

Tang J, Le S, Sun L, Yan X, Zhang M, Macleod J, Leroy B, Northrup N, Ellis A, Yeatman TJ et al. 2010. Copy number abnormalities in sporadic canine colorectal cancers. *Genome research* **20**(3): 341-350.

Tang J, Li Y, Lyon K, Camps J, Dalton S, Ried T, Zhao S. 2014. Cancer driver-passenger distinction via sporadic human and dog cancer comparison: a proof-of-principle study with colorectal cancer. *Oncogene* **33**(7): 814-822.

Vermorken JB, Herbst RS, Leon X, Amellal N, Baselga J. 2008. Overview of the efficacy of cetuximab in recurrent and/or metastatic squamous cell carcinoma of the head and neck

in patients who previously failed platinum-based therapies. *Cancer-Am Cancer Soc* **112**(12): 2710-2719.

Wei W, Bracher-Manecke JC, Zhao X, Davies NH, Zhou L, Ai R, Oliver L, Vallette F, Hendricks DT. 2013. Oncogenic but non-essential role of N-myc downstream regulated gene 1 in the progression of esophageal squamous cell carcinoma. *Cancer biology & therapy* **14**(2): 164-174.

Youmans L, Taylor C, Shin E, Harrell A, Ellis AE, Seguin B, Ji X, Zhao S. 2012. Frequent Alteration of the Tumor Suppressor Gene APC in Sporadic Canine Colorectal Tumors. *PloS one* **7**(12): e50813.

Zhang J, Ding L, Holmfeldt L, Wu G, Heatley SL, Payne-Turner D, Easton J, Chen X, Wang J, Rusch M et al. 2012. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**(7380): 157-163.

# CHAPTER 4

# CONCLUSIONS

By performing the first comprehensive genome-wide characterization of dog mammary and oral cancers, we found the strong molecular similarities between human and dog on these two cancers, and demonstrated dog cancer represents an excellent model of human cancer.

For dog mammary cancers, we studied 12 cases of the genomes, transcriptomes and epigenomes, including canine simple carcinomas and complex carcinomas, two major histologic subtypes of dog MC. The analysis reveals that canine simple carcinomas faithfully recapitulate many molecular features of human breast cancer. Notably, canine simple carcinomas closely cluster with basal-like human breast tumors in PAM50 classification, and thus could serve as a much-needed spontaneous cancer model for the basal-like subtype. Our analysis also indicates that canine complex cancers may arise from epigenomic rather genomic alterations.

One major subtype of dog mammary cancer, canine simple carcinomas, harbor extensive genomic aberrations and are molecularly homologous to human breast carcinomas. First, these strong molecular similarities make canine simple carcinomas as a great model in human breast cancer research. The dog mammary cancers, especially canine simple carcinomas, bridge a gap between traditional rodent models and human clinical trials. For example, for canine simple carcinomas with large CNAs, we could apply the dog-human comparison method for effective

driver-passenger discrimination. For drugs targeting the WNT or MAPK/ERK pathway, their efficacy, dosage, toxicity and schedule can be more accurately evaluated through clinical trials with dog patients, before entering human clinical trials. Second, canine simple carcinomas cluster with basal-like human tumors with an 82% chance in our PAM50 classification, indicating their highly similarities with human basal-like subtype. The further study with a larger sample size is needed to determine if canine simple carcinomas closely match the basal-like subtype. Compared to other subtypes, basal-like cancers are aggressive, have a poor prognosis and currently lack effective treatments. If confirmed, canine simple carcinoma could make significant contributions towards understanding and treating this worst subtype of human breast cancer.

The other major subtype of dog mammary cancer, canine simple carcinomas, with both luminal and myoepithelial cell proliferation, appear to originate from epigenomic rather than genomic alterations. Myoepithelial cell proliferation is rare in human breast cancer, so they receive far less attention than luminal cells and are understudied. However, myoepithelial cells are thought to be a part of the mammary stem cell niche, and possibly derive from stem cell or common progenitor based on our study. Dog mammary cancer, where myoepithelial cell proliferation is much more common, provides an ideal system to better understand the second major cell lineage of the mammary gland.

For dog oral cancers, we investigated 12 cases of spontaneous canine HNSCC (including 9 oSCC). Although the sample size is small, we found a strong dog-human molecular homology at various levels and reinforced the value of spontaneous dog cancers in oSCC basic and clinical

research. These include many of the same amplified/deleted or over/under-expressed genes, the same altered pathways and enriched functional groups, and similar types of large genomic changes and sequence mutations.

The strong dog-human molecular homology described in our study fundamentally justifies the use of our established dog-human comparison strategy for driver-passenger discrimination. Investigating both the dog and human oral cancers will reduce the number of driver candidate genes by at least half, and significantly reduce the downstream functional validation. Plus, dog oral cancers can serve as a great translational model, as our study indicates that they faithfully represent their human counterparts at the molecular level. Dog oral cancers can significantly accelerate the translation of basic research findings into successful clinical applications, coupled with less stringent FDA (Food and Drug Administration) regulations governing pet clinical trials. For example, for drugs targeting BRAF and other genes in the mitogenic pathway, their efficacy, dosage, toxicity and schedule can be more accurately evaluated through clinical trials with dog patients, before entering human clinical trials. This will significantly reduce the cost and accelerate new anticancer drug discovery.

These two projects are the pilot studies for dog mammary and oral cancers. Further studies with a larger sample size could detect the recurrently changed genes and pathways in dog mammary and oral cancers. Plus, further studies with more samples could be used for effective driver-passenger discrimination by the dog-human comparison method in the entire dog genome.