DISSERTATION

# USING THE DOG AS A MODEL TO INVESTIGATE ENVIRONMENTAL AND GENETIC RISK FACTORS FOR MATURE, ANTIGEN-DRIVEN LYMPHOPROLIFERATIVE DISORDERS

Submitted by

Julia Labadie

Department of Environmental and Radiological Health Sciences

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2017

Doctoral Committee:

Advisor: Sheryl Magzamen
Co-Advisor: Anne Avery

Brooke Anderson
Heather Feigelson
Paul Morley
Rodney Page

ABSTRACT

USING THE DOG AS A MODEL TO INVESTIGATE ENVIRONMENTAL AND GENETIC RISK FACTORS FOR MATURE, ANTIGEN-DRIVEN LYMPHOPROLIFERATIVE DISORDERS

Lymphoproliferative disorders are common hematopoietic neoplasms in both humans and dogs that are heterogeneous and have an uncertain, likely multifactorial, etiology. A subset of human lymphoproliferative disorders, including B cell chronic lymphocytic leukemia (B-CLL) and T zone lymphoma (TZL), are thought to be antigen-driven.[1-6] Dogs commonly develop these subtypes,[7, 8] allowing them to be used as natural models for the corresponding human diseases. Canine models are increasingly being used to ascertain genetic links to disease and conduct pre-clinical trials of novel therapies. In addition, dogs present a unique opportunity for identifying non-genetic risk factors for disease, as they spend most of their lives alongside their human counterparts, encountering many of the same environmental exposures. *Therefore, we hypothesized that human and canine lymphoproliferative disorders share risk factors, clinical presentation, and prognosis.* The goal of this dissertation was to contribute to the framework of canine cancer models and begin to elucidate novel ways we can use dogs to further our understanding of both mechanistic and epidemiologic aspects of human disease.

**Aim 1: Describe canine B-CLL presentation, clinicopathologic findings, and breed predisposition**. B-CLL is the second most common leukemia that affects humans in the United States.[9] First-degree relatives of B-CLL patients have an 8.5-fold higher risk for developing the disease.[10] However, a clear set of risk-associated genes has not been identified. Because of the

discrete genetic pools maintained by selective breeding, examining canine B-CLL could be a powerful way to study genetic risk factors for this disease. Little is known, however, about the population that develops canine B-CLL. We used submissions from the Colorado State University's Clinical Immunology laboratory to describe the typical patient presentation and associated clinical signs of canine B-CLL in addition to evaluating whether certain breeds have differential proportions of B-CLL. We determined canine B-CLL is typically a disease of older individuals with no sex predilection. Half of cases presented with peripheral lymphadenopathy or splenomegaly and one quarter had anemia. Small breed dogs and English Bulldogs appear to be predisposed to this disease, with English Bulldogs exhibiting a unique presentation that we hypothesize is associated with a more aggressive phenotype. Results from this aim helped set the foundation for the use of canine B-CLL as a model for its human counterpart, with the long-term goals of evaluating prognostic factors as well as identifying genetic risk factors for developing B-CLL.

**Aims 2 and 3: Use canine TZL to better understand the etiology and pathogenesis of human peripheral T cell lymphoma.** TZL, a rare indolent lymphoproliferative disorder in humans, appears to have a striking predilection for Golden Retrievers. Golden Retrievers comprise 12% of all dogs with lymphoproliferative disorders, but 40% of all dogs with TZL. This observation suggests a significant genetic risk factor(s) for this disease. Since TZL cells exhibit an activated phenotype, characterized by increased expression of CD21 and CD25,[6] we also hypothesized chronic immune stimulation was necessary for progression to overt disease.

We conducted a case-subclinical-control study of TZL among Golden Retrievers to investigate environmental and genetic risk factors for this disease. Owners of 508 Golden

Retrievers (140 cases, 221 subclinicals, 147 controls) aged 9 years or older participated in our study. Cases were recruited from submissions to Colorado State University's Clinical Immunology laboratory; flow cytometric diagnosis required identification of a homogenous expansion of T cells lacking expression of the cell surface marker CD45.[6] The reference population was recruited through 1) a database of owners interested in participating in research studies and 2) the submitting clinics of cases. These dogs were classified as controls or subclinicals based on flow cytometry.

**Aim 2: Identify genes associated with TZL in Golden Retrievers using a genome-wide association study**. A subset of 360 dogs (95 cases, 142 subclinicals, 101 controls) were genotyped using the Illumina CanineHD BeadChip.[11] Data were analyzed using genome-wide complex trait analysis software and significance was determined based on the 95% confidence intervals on the quantile-quantile plot. Associated regions were subsequently sequenced using a custom sequence capture array (NimbleGen SeqCap EZ Developer Kit) on an Illumina NextSeq 500. We found genome-wide significance in regions on chromosomes 8 and 14 that differentiated TZL cases from both subclinicals and controls. No statistically significant differences were noted between subclinicals and controls. Resequencing of the *chromosome 14* peak identified five missense mutations on three hyaluronidase genes clustered in the associated region. One variant of the SPAM1 gene was predicted to be damaging using PolyPhen-2.[12] Resequencing of the *chromosome 8* peak did not identify any coding mutations, but implicated potential modifying mutations for SEL1L, STON2, GTF2A1, CEP128, and TSHR. The majority of mutations were predicted to modify SEL1L, STON2, and CEP128, which have roles in vesicle trafficking and endoplasmic reticulum homeostasis: SEL1L encodes a protein for the

endoplasmic reticulum-associated protein degradation complex,[13] STON2 plays a role in receptor-mediated endocytosis,[14] and the role of CEP128 is not well characterized. Results from this aim identified a potential role of hyaluronan turnover in TZL pathogenesis. Additionally, mutations in SEL1L and STON2 may indicate accumulation of misfolded protein and/or changes in antigen presentation could influence development of TZL.

**Aim 3: Examine the association of chronic inflammatory conditions and TZL using a questionnaire-based case-control study.** Information about each dog's health history, signalment, lifestyle, and environmental exposures were obtained via an owner-completed questionnaire. Odds ratios and 95% confidence intervals were estimated using multivariable logistic regression. We identified several inflammatory risk factors with diverse mechanistic pathways were associated with an increased risk of TZL. Two of these, gastrointestinal disease and mange, were only associated with TZL cases, whereas bladder infection and eye disease were associated with both cases and subclinicals. We additionally identified two potentially protective factors, omega-3 supplementation and hypothyroidism. Our findings support the hypothesis that inflammation is a risk factor for TZL, and may indicate anti-inflammatory therapeutics could help treat or prevent this disease.

**Conclusions:** This dissertation contributed to the body of evidence that B-CLL and TZL are antigen-driven, and additionally suggested a genetic component of both diseases. Our findings can be used to inform future mechanistic studies to better understand disease pathogenesis and ultimately provide insights into novel treatments or preventive measures. On a broader scale, this dissertation demonstrated the utility of conducting epidemiologic studies

among canine populations, particularly among diseases that are rare in people but comparatively common in dogs.

# ACKNOWLEDGEMENTS

I feel extremely fortunate to be part of a field that is composed of such inquisitive, driven, and passionate people. At times, the process of obtaining a PhD can feel isolating, but all it takes is a quick look around to realize I am part of a large, interconnected community. I have so often depended on the support of my classmates, labmates, and advisors, and they have given without hesitation. It is with immense gratitude that I reflect back on all who helped me achieve this goal.

For the past five years, I have walked by the campus statue that reads, "If I have been able to see farther than others, it was because I stood on the shoulders of giants." I certainly do not feel that I have seen further than others, but any scientific success I had during this program has been in large part due to two small, but incredibly mighty giants. Drs. Anne Avery and Sheryl Magzamen, your drive, intellect, and passion for science never cease to challenge and inspire me. Thank you for the countless hours you put into my development as a researcher through your critical thoughts and sage advice.

To the rest of my committee (Drs. Brooke Anderson, Heather Feigelson, Paul Morley, Rod Page), thank you for helping to better my work and myself through your astute observations and questions. You have pushed me to think outside my comfort zone, expanded my knowledge base, and helped me focus when I felt overwhelmed by options. There are also countless other mentors and advisors who helped me along the way. In particular, Dr. Ginger Elvers at Uppsala University, whose good humor, extreme patience, and vast knowledge were essential to my completion of aim 2. I also must thank Dr. Susan Stover at UC Davis, whose

support, encouragement, and guidance was paramount in my decision to pursue a combined DVM/PhD degree.

To my DVM/PhD classmates, "people of epi," Avery labmates, and GAUSSI participants: thank you for always providing a critical review and fresh perspective of my research and presentations, for embracing my "differences," and allowing me to feel like part of each of your respective groups. Likewise, to the members of the Clinical Immunology laboratory, thank you for your humor while I stumbled around the lab and for your patience with calls, emails, and samples from countless Golden Retriever owners.

To the participants of the Golden Years Study, I am incredibly grateful for your contributions to my project, but even more so for the informal communications we had over the years. Your emails continually give meaning to my work; as each year goes by, I celebrate your dogs' birthdays and mourn your losses. I take every email to heart and know that the impact of this work reaches far beyond a scientific manuscript. May we one day reduce the burden of cancer among our beloved Golden Retrievers.

To my family, friends, and gym-mates: thank you for giving me shelter to escape from the demands of this program and forget about things for a while. Your love, encouragement, and sense of humor enabled me to regain perspective and focus on all the good around me. To my dad, you are with me always. Your whole-hearted devotion to all my childhood pursuits, from horseback riding to science fairs, empowered me to become the person I am today.

# DEDICATION

This dissertation is dedicated to the two most important “M’s” in my life:

My mother, for being my biggest supporter. At every step of the way you had my back. You celebrated my successes, shouldered my failures, and provided every form of encouragement in between. Thank you for your humility, tolerance, resilience, and occasional stubbornness.

Maggie, my faithful Golden Retriever, for being the best adventure buddy, science participant, and cuddle bug. Your unwavering happiness gives me a reason to smile every single morning. May you defy the odds and live a long, cancer-free life.

TABLE OF CONTENTS

# CHAPTER 1: INTRODUCTION

Non-Hodgkin's lymphoma (NHL) and leukemia are among the most common types of hematopoietic neoplasms in the United States, with over 130,000 new cases and nearly 45,000 deaths estimated in 2016.[15] Overall, NHL is estimated to be the 6th most common cancer based on new diagnoses and the 9th most common based on deaths, whereas leukemia is the 6th most common for deaths and 8th – 9th (for males and females, respectively) most common based on new diagnoses.[15] This discrepancy between prevalence and mortality highlights potential differences in disease aggressiveness, diagnosis, and treatment. NHL and lymphoid leukemias are also very common in dogs and are usually generalized into the disease category "lymphoma" or lymphoproliferative disorders (LPDs). While there are no current, population-based estimates of incidence or prevalence, canine lymphoma is considered the most common hematopoietic neoplasm in dogs.[16, 17] One of the most recent (from 1997-1998) estimates of annual incidence among United Kingdom dogs was 107 per 100,000 dogs for lymphoma and 11 per 100,000 dogs for leukemia.[18] Among people, the corresponding incidence rates are 19.5 per 100,000 and 13.5 per 100,000, respectively.[19] These estimates suggest that lymphoma is more common in dogs than people, highlighting the utility of the canine model for human lymphoma.

Lymphoma is a heterogeneous disease, with over 60 subtypes identified in humans.[20, 21] These subtypes vary both morphologically and behaviorally, leading to distinct clinical presentation, treatments, and prognoses. These subtypes can be split into two broad categories based on cell of origin (B and T lymphocytes). In humans, B cell neoplasms are much more

common, making up about 90% of cases in developed countries, and therefore more is known about their etiology and treatment.[21] Within both B and T cell neoplasms, subtypes are categorized as precursor acute lymphoblastic leukemia/lymphoma or mature types, with the majority for each being mature.[21] The World Health Organization classification includes 17 subtypes of peripheral B cell neoplasms and 18 subtypes of peripheral T cell neoplasms (Table 1.1). Among B cell neoplasms, diffuse large B cell lymphoma (DLBCL) and follicular lymphoma (FL) are the most common, each accounting for about 25% of all lymphomas in developed regions.[21, 22] T cell lymphomas are more rare, with the largest category being peripheral T cell lymphoma (PTCL) not otherwise specified, accounting for less than 3% of all lymphomas.[20, 21]

Lymphoma can also be categorized based on prognosis into aggressive and indolent (slow-growing) types (Table 1.1). Aggressive types are generally more responsive to modern treatments (combination chemotherapy); while overall 5-year survival is around 60%, more than 50% of patients experience complete remission.[23] In contrast, the median survival for patients with indolent lymphoma is about 20 years, but patients may experience continuous relapses, requiring multiple rounds of chemotherapy.[24] At early stages, some types of indolent lymphoma may not require treatment and clinicians recommend "watchful waiting".[25]

**Table 1.1. World Health Organization classification of human NHL.**[20, 21]

| Cell of origin | | Behavior | Subtype |
|---|---|---|---|
| B cell | **Precursor** | Aggressive | • B acute lymphoblastic leukemia/lymphoblastic lymphoma (B-ALL) |
| | **Peripheral** | Aggressive | • Burkitt lymphoma (BL)<br>• Diffuse large B cell lymphoma (DLBCL)<br>• Follicular lymphoma, high grade (FL)<br>• Mantle cell lymphoma (MCL) |
| | | Indolent | • B cell chronic lymphocytic leukemia/small lymphocytic lymphoma (B-CLL)<br>• Follicular lymphoma, low grade (FL)<br>• Hairy cell leukemia (HCL)<br>• Lymphoplasmacytic lymphoma/immunocytoma (LPL)<br>• Marginal zone lymphoma (MZL)<br>  ○ Splenic, nodal, or extranodal (Mucosa-associated lymphatic tissue)<br>• Plasmacytoma/plasma cell myeloma |
| T/NK cell | **Precursor** | Aggressive | • T acute lymphoblastic leukemia/lymphoblastic lymphoma (T-ALL) |
| | **Peripheral** | Aggressive | • Adult T cell lymphoma (human T-lymphotrophic virus 1+)<br>• Aggressive NK cell leukemia<br>• Anaplastic large cell lymphoma, primary systemic type<br>• Angioimmunoblastic T cell lymphoma (AITL)<br>• Extranodal T-/NK-cell lymphoma, nasal type<br>• Hepatosplenic gamma/delta T cell lymphoma<br>• Peripheral T cell lymphoma (PTCL)<br>• Subcutaneous panniculitis-like T cell lymphoma |
| | | Indolent | • Anaplastic large cell lymphoma, primary cutaneous type<br>• Mycosis fungoides/Sezary syndrome (MF)<br>• T cell chronic lymphocytic leukemia/prolymphocytic leukemia (T-CLL)<br>• T cell granular lymphocytic leukemia |

**Canine lymphoma**

Canine lymphoma shares many of the same subtypes as human lymphoma, and can be categorized using similar criteria. In a 2011 study by Valli et al.,[7] the World Health Organization human classification system was applied to 300 biopsies of canine lymphoma. Like human lymphoma, DLBCL was the most common subtype, but it accounted for a larger proportion of cases than among humans (48% vs. 25%). Marginal zone lymphoma (MZL) was the second most

common B cell lymphoma, accounting for about 4% of all cases. Interestingly, FL appears rare in dogs, whereas it is quite common in humans.[7, 21] Proportionally, T cell lymphomas were much more common in dogs than people, with PTCL accounting for over 25% of cases and T lymphoblastic lymphoma accounting for approximately 4% of cases (vs. ~1% in humans[21]). Almost half of the PTCLs were T zone lymphoma (TZL), a subtype that is extremely rare in humans. Figure 1.1 below compares the relative distribution of selected subtypes of lymphoma, comparing canine data from our laboratory to human data from a recent publication.[21] This figure highlights the differences in the relative distribution of lymphoma subtypes for dogs and humans. These differences highlight the utility of comparative models to understand why certain subtypes are more common in one species. Additionally, because T cell lymphomas are more common among dogs, they provide an excellent opportunity to conduct studies where it would be difficult to gather enough human patients.

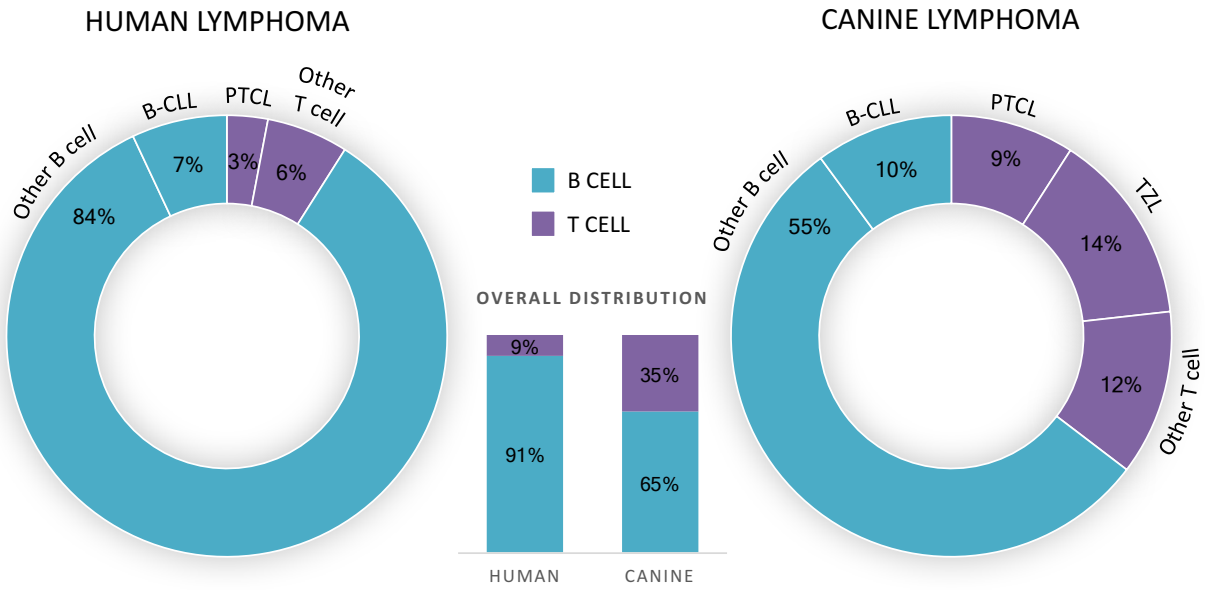



**Figure 1.1. Relative distribution of selected subtypes of canine lymphoma and human NHL.**[21] PTCL, peripheral T cell lymphoma; TZL, T-zone lymphoma; B-CLL, B-cell chronic lymphocytic leukemia.

Canine lymphoma has a similar presentation and prognosis to human lymphoma and is therefore treated with many of the same drugs. Combination chemotherapy is standard treatment for dogs with aggressive lymphoma, resulting in a median remission time of 7-9 months and median overall survival of 10-13 months.[26, 27] As with human lymphoma, indolent types offer a much better prognosis, with median survival exceeding 19 months.[8, 28]

**The dog as a model for human cancer**

The canine model has been accepted for many human cancers, including lymphoma, due to strong similarities histologically as well as clinically.[29] The utility of the canine model is wide-ranging, with benefits for clinical, mechanistic, genetic and epidemiologic studies across many types of cancers. Clinically, there is good infrastructure for clinical trials in dogs, especially among veterinary teaching hospitals that see large caseloads and are accustomed to conducting research.[29, 30] Since there is no gold-standard treatment for canine lymphoma, new drugs can be tried as first-line treatment as opposed to last-resort in human clinical trials.[29] Many cancer drugs are currently undergoing clinical trials in dogs (vetcancertrials.org), and some have already moved forward into human trials. One example of this is PAC-1, which shows promise in treatment of brain cancer.[31]

Since canine lymphoma occurs spontaneously, as opposed to experimentally-induced in rodent models, it is believed to better represent the "natural" cancer system seen in human patients, which is more heterogeneous than tumor models in rodents.[29] This is helpful from a mechanistic perspective because the same factors are at play as with a spontaneously-occurring human cancer, including diet, environment, and host defenses (e.g. immune system). From an

epidemiologic perspective, dogs share much of the same environment as people, live in the same households and experience similar lifestyle factors. Since dogs have considerably shorter lives than people, it is more feasible to conduct cohort studies to assess risk factors for lymphoma as well as to conduct follow-up studies to assess survival and related issues. Many follow-up studies have been conducted to evaluate survival and prognostic factors, particularly for cancer. However, no large-scale observational cohorts had been undertaken until the Golden Retriever Lifetime Study began in 2012.[32] This groundbreaking study will follow 3,000 Golden Retrievers throughout their lives to identify risk factors for cancer and other diseases.

**Genetic analyses in dogs**

As most mammals share a significant proportion of their DNA, canine genetic research can also be translated to human genetic research. Human genomes contain about 3.3 billion base pairs which constitute 23 pairs of chromosomes, whereas canine genomes are smaller (2.8 billion base pairs) and divided over 38 pairs of chromosomes. However, the overall similarity is high, with dog alignments covering about 80% of human transcripts and 75% of all genes.[33] From an evolutionary perspective, the development of dog breeds was relatively recent. A consequence of the selective breeding and population bottlenecks is that many dog breeds have a higher prevalence of specific diseases,[34, 35] suggesting a limited number of loci are responsible for those diseases.[36] When DNA recombines, DNA segments, called haplotype blocks, are moved from one chromosome to another (Figure 1.2). If this recombination is recent, a larger haplotype block will be shared among related individuals. As time goes on, more recombination will happen, breaking these blocks apart, resulting in shorter shared regions.

Within breeds, there is reduced genetic variation, evidenced by large haplotype blocks (500kb – 1Mb).[37-39] However, the diversity is much greater across breeds, with the average haplotype block around 10kb, comparable to what is seen in humans.[38] Portions of some haplotypes are conserved across breeds, suggesting genetic risk factors may be shared across breeds.[38]

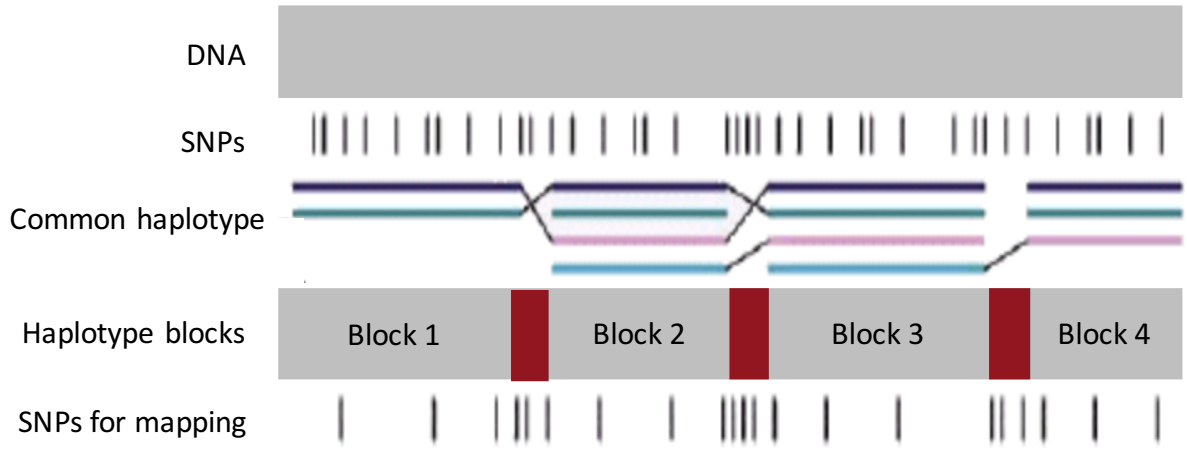



**Figure 1.2. Depiction of how haplotype blocks relate to SNP selection for GWAS chips.**

The larger haplotype blocks seen within breeds are beneficial from an analytical standpoint. Researchers have compared the genomes of multiple dog breeds to find the locations of common variants, called single nucleotide polymorphisms (SNPs), that are present in at least 1% of the population. This work has led to the discovery of millions of SNPs; however, because SNPs within a haplotype block are inherited together, genotyping only a few SNPs within that block can provide a representation of risk for that block. Because haplotype blocks are so large in dogs, the number of SNPs that need to be genotyped for a genome-wide association study (GWAS) can be substantially reduced. Human GWAS SNP microarrays typically contain >500,000 SNPs to have an "adequate" representation of the human genome, whereas the standard canine SNP microarray contains only 170,000 SNPs.[11] In addition, as

GWAS analyses involve multiple comparisons, there is always a concern that multiple testing leads to many false positive findings; fewer SNPs correspond to fewer comparisons and thus less stringent p-value adjustments are needed.

If we look within one breed of dog, this reduced genetic variation also means we are assessing a more homogeneous population. From a statistical perspective, this reduces variability, or "noise", and enhances our power to detect genetic differences associated with our outcome of interest.[40] Thus, we can use smaller sample sizes in canine vs. human GWAS.[37, 38] There is no universal sample size calculator for canine GWAS. The appropriate sample size is a balance of homogeneity of the population (i.e. level of inbreeding), whether the trait of interest is mono- or multigenic, and the expressivity and penetrance of the loci (i.e. the extent to which the trait is expressed and the percent of affected individuals with the associated allele). Using simulated data, one researcher estimated that, for a study evaluating a multigenic trait using 15,000 SNPs, a sample size of 100 affected and 100 unaffected dogs would give 97% power to detect an five-fold increase in disease risk, but only 50% power to detect an 2-fold increase in disease risk.[38]

However, the genetic homogeneity also means more careful selection of study participants is needed to ensure close relatives (i.e. closer than the grandparent level)[41] are not included in the study. Ideally, we would ensure this by conducting a pedigree analysis before enrollment to avoid the cost of genotyping dogs that would later be removed through relatedness filtering. Even within a breed, population stratification can exist due to founder effects with popular sires.[40] If not accounted for, systematic differences between cases and controls can drive the genetic differences observed, thereby biasing our effect estimates. Thus,

after removing overly related individuals it may still be necessary to adjust for residual confounding due to underlying population stratification. This is a known problem among Golden Retrievers of American vs. European descent, which, according to one publication,[42] are roughly as diverged as European and Asian human populations. Thus, care must be taken to select dogs from within one geographic region. Researchers generally recommend using a single dog breed for GWAS to ensure appropriate identification and control of population stratification.[42, 43]

Certain aspects of quality control that are standard for human GWAS[44] should be reconsidered when conducting canine GWAS. First, the canine genome contains extensive regions of homozygosity that are interspersed with highly heterozygous regions.[38] As a result, filtering based on heterozygosity rate, which is commonly done for human GWAS quality control, is not appropriate for canine GWAS. Second, because canine populations are highly selected, Hardy-Weinberg equilibrium is violated and SNPs that deviate from Hardy-Weinberg equilibrium should not be filtered. Lastly, while filtering for low minor allele frequency is still appropriate for canine GWAS, a much larger fraction of SNPs will be removed. Because canine GWAS SNP chips are designed to be used for multiple breeds, a subset of SNPs will not be polymorphic if we look within one breed. When the CanineHD BeadChip was tested on 26 diverse dog breeds, the number of polymorphic loci (minor allele frequency $\geq$0.05) ranged from about 85,000 to over 126,000 (of >170,000 SNPs).[11] Thus, depending on the breed being studied, up to 50% of SNPs could be filtered out due to low or fixed minor allele frequency. With the exception of these considerations, canine and human GWAS are carried out in the same way and the same tools can be used for quality control and analysis. Certain software, such as

PLINK[45] and GCTA,[46] require the use of a flag to indicate there will be more chromosomes than would be expected for a human genome.

**Observational risk factor epidemiology for canine lymphoma**

*Study population*

Dogs have largely been untapped for observational risk factor epidemiology. Currently, no population-based canine cancer registry exists, although multiple attempts have been made to create such a registry (reviewed in Bronden et al., 2007[47]). The Veterinary Medical Database (https://vmdb.org/), which began in 1964, was arguably the most successful attempt. This database contains records from 26 participating veterinary teaching hospitals, geographically dispersed in the United States and Canada. It has been the basis of multiple hospital-based case-control studies for canine risk factor epidemiology, particularly in its earlier years.[48-50] For those studies, this database provided a large population of dogs with confirmed diagnoses. However, submission to this database has not been consistent over the years and currently less than eight hospitals submit data;[47] thus the representativeness and completeness of this database is uncertain.

Studies that have not used the Veterinary Medical Database have largely been hospital-based, and almost all have used teaching hospitals. Since case recruitment often occurs over an extended period of time, controls are frequently matched by year of diagnosis. Selection of a hospital-based control population has been variable, with choices including: 1) all non-lymphoma tumor diagnoses,[48, 51] 2) all non-cancer diagnoses,[48] 3) benign tumor diagnoses,[52] 4) non-cancer chronic diseases,[52] and 5) all non-lymphoma diagnoses.[49, 53, 54] Some studies have also

used multiple control groups to assess the impact each group had on their results.[48, 52] Controls have also variably been matched by sex, age, and/or location of residence.

In studies conducted outside the U.S., various other study populations have been used, including pet insurance registries and tumor registries. However, these studies have only assessed age, sex, and breed as risk factors.[55, 56]

In general, hospital-based studies are cautioned in epidemiology due to concerns for Berkson's bias.[57, 58] However, in veterinary epidemiology, where our case source is almost always obtained through veterinary clinics or diagnostic laboratories, hospitals may be our most representative source of controls, as many dogs will not ultimately be diagnosed or treated by veterinarians. As such, population-based controls likely include a subpopulation of dogs that would not have been diagnosed at hospitals or included in the case population. A potential exception is the use of the Canine Lifetime Health Project database (described in chapter 4) or breed registries, which may represent owners of a higher socioeconomic status who are more likely to go to the veterinarian.

*Exposure assessment*

Exposure assessment for observational risk factor studies of canine lymphoma has been variable. Two hospital-based studies[48, 52] used mailed questionnaires to assess exposure to household chemicals, particularly lawn chemicals (e.g. herbicides and pesticides). Response rate was low (40%), but one study included telephone interview follow-up,[48] which increased the overall response rate to 80%. This study, however, received considerable criticism about

potential recall bias and analytic strategy,[59] prompting a subsequent publication with a reanalysis.[60]

Two other studies focused on environmental exposures: one assessed exposure through both phone interviews and household visits (measured electromagnetic fields);[51] the other used an ecologic approach, assessing proximity to hazardous waste disposal sites.[53] Interestingly, this study also utilized a standardized intake questionnaire that all owners complete at the time of their hospital visit to evaluate the dog's diet, source of drinking water, exposure to passive tobacco smoke, use of antiparasitics, and exposure to herbicides and electromagnetic fields.[53]

Only one study assessed biologic samples from study participants. This study received serum samples from dogs to conduct laboratory tests evaluating potential infection with specific vector-borne pathogens.[54] Studies that have assessed signalment[50, 55, 56, 61] (i.e. sex, breed, age) or specific health-related factors (i.e. immune-mediated disease[49]) have used medical records and thus avoided interacting with owners for exposure ascertainment.

Overall, the field of lymphoma observational risk factor epidemiology has potential for improving our understanding of both canine and human cancers. Few studies have undertaken this goal, and thus there is much room to expand our current knowledge. As with all questionnaire-based research, there is the concern for recall bias when asking about previous exposures. Dogs may be beneficial from this standpoint, as their lifetime is substantially shorter than people, and thus owners do not have to recall as many years. However, as not all dogs live with the same owner throughout their life, recall may be limited by the duration of ownership and certain histories may be unknown. Important considerations and potential sources of bias when designing canine case-control studies will become more apparent as this field grows.

**Risk factors for human lymphoma**

The pathogenesis of most forms of lymphoma is uncertain and likely multifactorial. Research on lymphoma risk factors has been limited and generally focused on overall risk factors as opposed to subtype-specific risk factors. Risk factors that have been identified include: 1) age, with older age conferring higher risk; 2) gender, with males at higher risk than females for most subtypes; 3) race, ethnicity, and geography, with NHL generally being a disease of Whites in developed countries; 4) certain chemical exposures, including benzene, insecticides, and herbicides; 5) altered immune function, including higher risk among people who receive immune suppressants and those with autoimmune disorders; 6) certain infectious agents, including known lymphocyte-transforming viruses (e.g. human T cell lymphoma virus), infections that cause chronic immune stimulation (e.g. Helicobacter pylori), and infections that cause immunosuppression (e.g. human immunodeficiency virus); and 7) genetic variations for some subtypes.[62] Many of these risk factors appear to be unique to specific subtypes of NHL, highlighting the importance of studying subtypes independently. However, some subtypes, especially T cell subtypes, are relatively rare in people, making it difficult to ascertain enough cases for well-powered epidemiologic studies.

Recently, the International Lymphoma Epidemiology Consortium (InterLymph) was created in an effort to pool studies to better assess subtype-specific risk factors. One study by Morton et al.[63] assessed family history, medical history, lifestyle, and occupational risk factors for multiple lymphoma subtypes, including acute lymphoblastic leukemia (ALL; n=152), Burkitt's lymphoma (BL; n=295), chronic lymphocytic leukemia (CLL; n=2,440), DLBCL (n=4,667), FL (n=3,530), hairy cell leukemia (HCL; n=154), lymphoplasmacytic lymphoma (LPL;

n=374), mantle cell lymphoma (MCL; n=557), MZL (n=1,052), mycosis fungoides (MF; n=324), and PTCL (n=584) compared to 23,096 controls (Table 1.2). Hematologic malignancy in a *first-degree relative* was significantly associated with all subtypes except BL and MF, suggesting a substantial genetic risk for most subtypes. *Medical histories* assessed included autoimmune disease, hepatitis C virus seropositivity, and atopic disease. History of a B cell activating disease was significantly associated with three B cell subtypes: MZL, LPL, and DLBCL. Specifically, both Sjögren's syndrome and systemic lupus erythematosus were associated with MZL and LPL; systemic lupus erythematosus was also associated with MF and PTCL. History of a T cell activating disease was significantly associated with both T cell subtypes (MF and PTCL), with celiac disease associated with PTCL and systemic sclerosis associated with MF, BL, and HCL. Hepatitis C virus seropositivity was associated with an increased risk of MZL, BL, LPL, and CLL. Interestingly, both hay fever and allergy decreased the risk of multiple subtypes, including BL, DLBCL, FL, MCL, and ALL; allergy also decreased the risk for PTCL and CLL. *Lifestyle factors* assessed included anthropometric factors, alcohol consumption, cigarette smoking, recreational sun exposure, and socioeconomic status. Increased body mass index as a young adult and height were significantly associated with an increased risk of DLBCL, FL, HCL, and ALL; height was additionally associated with PTCL, BL, CLL, and MCL. Alcohol consumption of at least 1 drink per month was associated with a decreased risk of MF, PTCL, MZL, BL, and DLBCL. Duration of cigarette smoking increased the risk of all subtypes except BL, DLBCL, and CLL, whereas recreational sun exposure decreased the risk of all subtypes except MF, LPL, HCL, and ALL. Lastly, *occupational history* was assessed. Teaching was associated with a significantly decreased risk of MZL, BL, and LPL, whereas being a painter

was associated with a significantly increased risk of MF and BL, and being a general farm worker was associated with a significantly increased risk of MF, BL, CLL, and MCL. Hierarchical clustering was conducted to identify subtypes with similar risk profiles, and determined that B and T cell subtypes clustered into separate nodes. Within the B cell subtypes, there was more heterogeneity, with three nodes distinguishing B cell subtypes.

**Table 1.2. Summary of selected results from Morton et al**.[63] Red "I" indicates significantly increased risk, Blue "D" indicates significantly decreased risk, and "-" indicates no statistically significant association. Mycosis fungoides (MF), peripheral T cell lymphoma (PTCL), marginal zone lymphoma (MZL), Burkitt lymphoma (BL), lymphoplasmacytic lymphoma (LPL), diffuse large B cell lymphoma (DLBCL), chronic lymphocytic leukemia/small lymphocytic leukemia (CLL/SLL), follicular lymphoma (FL), mantle cell lymphoma (MCL), hairy cell leukemia (HCL), acute lymphoblastic leukemia (ALL).

| Exposure Category | Specific Exposure | MF | PTCL | MZL | BL | LPL | DLBCL | CLL/SLL | FL | MCL | HCL | ALL |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Family history of hematologic malignancy | | - | I | I | - | I | I | I | I | I | I | I |
| Autoimmune disease | Any B cell activating disease | - | - | I | - | I | I | - | - | - | - | - |
| | Sjogren's syndrome | - | - | I | - | I | - | - | - | - | - | - |
| | Systemic lupus erythematosus | I | I | I | - | I | - | - | - | - | - | - |
| | Any T cell activating disease | I | I | - | - | - | - | - | - | - | - | - |
| | Celiac disease | - | I | - | - | - | - | - | - | - | - | - |
| | Systemic sclerosis/scleroderma | I | - | - | I | - | - | - | - | I | - | - |
| Hepatitis C virus seropositivity | | - | - | I | I | I | I | - | - | - | - | - |
| Atopic disease | Hay fever | - | - | - | D | D | D | - | D | D | - | D |
| | Eczema | I | - | - | - | - | - | - | - | - | - | - |
| | Allergy | - | D | - | D | - | D | D | D | D | - | D |
| Anthropometric factors | Body mass index as a young adult | - | - | - | - | - | I | - | I | - | I | I |
| | Height | - | I | - | I | - | I | I | I | I | I | I |
| Alcohol consumption ( ≥1 drink per month) | | D | D | D | D | - | D | - | - | - | - | - |
| Cigarette smoking (duration of smoking) | | I | I | I | - | I | - | - | I | I | - | I |
| Recreational sun exposure | | - | D | D | D | - | D | D | D | D | - | - |
| Socioeconomic status | | D | D | - | D | - | D | - | - | D | - | D |
| Occupational history | Teacher | - | - | D | D | D | - | - | - | - | - | - |
| | Painter | I | - | - | I | - | - | - | - | - | - | - |
| | General farm worker | I | - | - | I | - | - | I | - | - | I | - |

The Morton et al. study suggested the importance of immune dysregulation for certain lymphoma subtypes, including PTCL, MZL, BL, DLBCL, and LPL.[63] The finding that B and T

cell activating diseases were associated with B and T cell lymphoma subtypes, respectively, supports this observation. This is corroborated by a previous study linking autoimmune diseases to PTCL, MZL, and DLBCL.[64] The mechanism behind this is believed to be related to chronic antigen stimulation, with the highest lymphoma risk seen in patients with most severe inflammatory disease.[64, 65] Additionally, other studies have found immunosuppression via organ transplantation or human immunodeficiency virus infection is a risk factor for DLBCL, LPL, MZL, and PTCL.[66-71] While CLL did not follow the same pattern in these studies, a prior study has shown farming, which may be related to chronic immune stimulation, is associated with CLL risk.[72] Additionally, laboratory-based studies have highlighted the importance of chronic antigen stimulation in the pathogenesis of CLL.[1-4] Overall, these studies suggest an important role of the immune system in the pathogenesis of lymphoma, especially among specific subtypes.

While a major strength of InterLymph is the increased sample sizes of lymphoma subtypes, some subtypes and exposures still suffered from low sample sizes, resulting in lack of power. Additionally, methodological problems can arise from pooling studies. Because the studies were designed independently, some exposure variables had to be re-categorized to be pooled, leading to loss of detail. In addition, only exposures that were measured in all studies could be evaluated, so only a subset of exposures of interest was included in the study. To improve exposure ascertainment, future multi-center studies would need to be designed *a priori* to have all participants complete the same, comprehensive questionnaire.

**Risk factors for canine lymphoma**

Few studies have assessed risk factors for canine lymphoma. Of those that have, history of immune-mediated thrombocytopenia,[49] exposure to 2,4-dichlorophenoxyacetic acid herbicides,[48] and exposure to electromagnetic fields[51] were found to be risk factors for canine lymphoma as a whole. Other studies have assessed a potential breed predisposition for canine lymphomas as a surrogate for a potential genetic risk factor.[55, 73, 74] Data from our lab and other publications suggest that certain subtypes are more likely to have a genetic component, as evidenced by skewed breed distributions, whereas others are more likely to have a lifestyle or environmental component, as evidenced by a more evenly-distributed breed distribution.[55, 73, 74]

To date, few studies have been conducted assessing genetics of canine lymphoma. Recently, a GWAS of B cell lymphoma and hemangiosarcoma[75] identified two potential shared predisposing loci on chromosome 5. However, whole genome resequencing did not identify any coding changes associated with the risk haplotypes. Another recent study[76] compared tumor whole exome across two breeds with T cell lymphoma (Golden Retrievers and Boxers) and two breeds with B cell lymphoma (Golden Retrievers and Cocker Spaniels). The study found similarities between the B cell lymphomas, including recurrent mutations in TRAF3-MAP3K14, FBXW7, and POT1. However, there was little overlap in the T cell lymphomas, with Boxers generally having mutations in the PTEN-mTOR pathway whereas Golden Retrievers had mutations in cellular metabolism genes. While these studies differentiated B and T cell lymphoma, they did not separate lymphomas by subtype. In Golden Retrievers, two subtypes of T cell lymphoma are common, one of which follows an aggressive clinical course, and the other (TZL) is indolent. Boxers more commonly develop the aggressive T cell type and rarely

get TZL. It is likely that the genetic basis of aggressive and indolent T cell lymphomas differs, so considering them together would attenuate effect estimates. A similar problem exists with B cell lymphomas, which is typically DLBCL, a more aggressive disease, or B-CLL, a more indolent disease. Cocker Spaniels get both diseases, whereas Golden Retrievers generally get DLBCL. However, since B-CLL typically presents in the blood and this study used lymph node samples for diagnosis, it is likely that the majority of tumors analyzed were DLBCL, explaining the more consistent findings for B cell lymphomas vs. T cell lymphomas.

For this dissertation, we focus on two subtypes that we believe have a genetic and environmental component, B-CLL and TZL. These subtypes are both indolent and arise from mature, antigen-driven lymphocytes,[1-5] suggesting an environmental stimulus. Additionally, data from our lab also suggests a breed predisposition, leading us to believe genetics play a role as well. By focusing on specific subtypes of lymphoma, we can leverage the benefits of a canine model to better understand risk factors that have been difficult to study in human populations. Since canine and human lymphomas are similar morphologically and behaviorally, it follows that they may share risk factors, clinical presentation, and prognosis. **Thus, the overarching goal of this dissertation is to establish both canine B-CLL and TZL as models for their human counterparts and to evaluate risk factors for the development of these diseases**.

**B cell chronic lymphocytic leukemia**

*Overview of human B cell chronic lymphocytic leukemia*

B-CLL is the second most common leukemia affecting people in the United States, accounting for almost 20,000 new cases annually.[15] In humans, CLL is a B cell neoplasm that

primarily affects older adults (average age at diagnosis ~72 years) and has an indolent course, with about 79% of patients surviving 5 years.[77] Patients generally present with vague symptoms including fatigue, weight loss, or fever, and the diagnosis is often an incidental finding during routine bloodwork.[15, 78] B-CLL is diagnosed by the presence of greater than 5 x $10^9$/L clonal small, mature B cells that express low levels of surface immunoglobulin and high levels of CD5.[79] Cytologically, these cells appear as small, mature lymphocytes with clumped chromatin and scanty cytoplasm.[79] There are two main systems for staging B-CLL, Rai and Binet, that take into account organ involvement and blood abnormalities.[79] The Rai staging system classifies patients as low-risk (stage 0) if only a lymphocytosis is present, intermediate-risk (stage I/II) if lymphadenopathy, splenomegaly, or hepatomegaly are present without any other blood abnormalities (besides lymphocytosis), and high-risk (stage III/IV) if anemia and/or thrombocytopenia are present with or without organ involvement. While B-CLL generally progresses slowly, overall survival can differ strikingly depending on immunoglobulin heavy chain variable region genes ($IgV_H$) mutation status.[79]

*Risk factors for human B cell chronic lymphocytic leukemia*

Risk factors for B-CLL have not been clearly defined, but a few findings have been consistent. First-degree relatives of B-CLL patients have an 8.5-fold higher risk for developing B-CLL,[10] suggesting a genetic risk factor. A potential role of genetics is also highlighted by the geographic differences in B-CLL incidence, with very low incidence in Asian countries that remains low with migration to the United States.[80] B-CLL incidence appears highest among people of European descent, with risk increasing with age and consistently higher among

males.[15, 81] Several GWAS have sought to identify the specific genetic risk factor involved in B-CLL.[82-86] These studies have found multiple loci that individually confer modest risk, but together may be responsible for a larger increase in risk. The genes associated with these SNPs are involved in regulation of transcription as well as B cell development, differentiation, and apoptosis. In addition to the genetic component of B-CLL etiology, the importance of $IgV_H$ mutation status, combined with restricted $IgV_H$ gene use, suggests a role of chronic antigen stimulation in the pathogenesis of B-CLL.[1-4] Other studies have indicated a history of autoimmune disorders[87] and certain occupations (e.g. farming)[72] increase B-CLL risk, further supporting the role of immune stimulation. Together, this suggests both a genetic and environmental role in B-CLL pathogenesis.

*Overview of canine B cell chronic lymphocytic leukemia*

What is known about canine B-CLL appears to parallel that of human B-CLL. Canine B-CLL is generally a disease of older dogs (average age at diagnosis ~10 years) and has an indolent clinical course, with most dogs still alive one year after diagnosis.[88] Dogs are also often diagnosed incidentally; if present, symptoms are vague and include lethargy and anorexia.[88-90] There are no consensus criteria for the diagnosis of canine B-CLL; unlike human B cell neoplasms, canine B cell neoplasms do not express the antigen CD5, so this cannot be used to distinguish B-CLL from other B cell lymphoproliferative disorders. Therefore, the diagnosis has typically relied on the morphologic characterization of cells described as small, mature and with clumped chromatin, with exclusion criteria including expression of CD34 (a marker of immature cells), or the presence of greater than 30% lymphoblasts[91,92] A prior study from our

laboratory[93] showed that, independent of morphologic criteria, B cell leukemias classified as "small" by flow cytometry of the peripheral blood have a median survival of greater than 1,000 days, consistent with a diagnosis of B-CLL. Because of this, we favor flow cytometric diagnosis of at least 5 x $10^9$/L lymphocytes, characterized as a homogeneous expansion (>60%) of "small" CD21+ lymphocytes. In flow cytometry, the forward scatter parameter is a measure of overall cell size. The classification of "small" is based on the ratio of the geometric mean of forward scatter of B cells to neutrophils being less than 0.55. There has been no systematic description of prognostic factors or staging for canine B-CLL, but we expect them to be similar to human B-CLL because of the overall similarity between canine and human lymphoma.[40]

*Risk factors for canine B cell chronic lymphocytic leukemia*

No risk factors have been established for B-CLL in dogs. Because of the strong familial risk seen in human B-CLL[10, 94, 95] and recent GWAS findings,[82-86] we suspect a genetic risk factor is also important in canine B-CLL. In addition, recent data from our laboratory suggests $IgV_H$ gene use is also restricted in canine B-CLL (Rout et al., manuscript in preparation), providing evidence that chronic antigen stimulation plays a role in canine B-CLL pathogenesis as well. Much insight can be gained by evaluating the utility of canine B-CLL as a model for human B-CLL. While distinct subtypes of human B-CLL have been identified, it is unknown whether they are tied to specific genetic risk factors and it can be difficult to ascertain enough cases to study the subtypes independently. By identifying breeds with distinct presentations of B-CLL, we can target cases within those breeds and thus may be better able to assess genetic risk for those subtypes.

**T zone lymphoma**

*Overview of human peripheral T cell lymphomas*

Among people, T cell lymphomas are relatively rare, accounting for about 10% of lymphomas in the United States.[20, 21, 96] These diseases are loosely grouped into PTCL and aggressive NK/T cell lymphoma. Among PTCLs, the most common subtype is PTCL-not otherwise specified (PTCL-NOS), which accounts for about 30% of all PTCLs.[20, 21] As the name suggests, this is a heterogeneous category that includes both nodal and extranodal mature T cell lymphomas. The average age at diagnosis of PTCL-NOS is 60 years, with a higher prevalence in males than females (66% vs. 34%).[97] PTCL-NOS is generally diagnosed at later stages, with 69% of individuals diagnosed at stage III or IV.[97] Overall, PTCL is believed to have a worse prognosis than aggressive B cell lymphomas,[98, 99] with a 5-year overall survival of 32%.[97] As with aggressive B cell lymphomas, treatment is generally multi-agent chemotherapy; while response rates are relatively high (60%), frequent relapses lead to the low overall survival.[97, 98, 100] However, the heterogeneity of PTCL and rarity of individual subtypes makes it difficult to determine whether specific subtypes have better prognoses.

*Risk factors for human peripheral T cell lymphomas*

This heterogeneity and rarity has also made mechanistic studies difficult, impeding our ability to identify modifiable risk factors and develop targeted treatments. Some types of PTCL have known viral etiologies, such as the association of human T cell lymphotrophic virus-1 and adult T cell leukemia/lymphoma[101] as well as the association of Epstein-Barr virus and nasal type T cell lymphoma.[102] Using pooled case-control studies, the InterLymph project identified

family history, medical, lifestyle, and occupational risk factors for PTCL.[103] Celiac disease, eczema, and psoriasis are associated with increased PTCL risk, while allergy was associated with a decreased risk. This is believed to be due to differences in immune responses, with atopic stimuli promoting type 2 helper T cell (Th2) responses and inflammatory and autoimmune conditions promoting type 1 helper T cell (Th1) responses. However, a major limitation of this study was the use of all PTCLs, a heterogeneous outcome. PTCL subtypes generally have different cells of origin, representing differing etiologies and thus distinct risk factors. By pooling the subtypes together, researchers risk attenuation of the effects of individual risk factors. Additionally, pooled analyses are subject to survival bias because more aggressive subtypes will be underrepresented. These limitations, however, are difficult to avoid due to the rarity of PTCL subtypes; it is extremely difficult to gather enough cases for epidemiologic studies, even with use of InterLymph, which pooled 15 studies. Canine research is an excellent way to avoid these limitations because T cell lymphomas are more common among dogs, allowing us to study subtype-specific risk factors.

As gene-expression profiling has become more readily available, progress has been made in subtyping of PTCL based on cell of origin.[104, 105] Angioimmunoblastic T cell lymphoma is now recognized to arise from T follicular helper cells, aptly named for their role in helping follicular B cells form germinal centers.[105] This has helped explain the B cell symptoms commonly seen with this disease. Cell of origin is also tightly tied to etiology. For instance, since Th1 cells are involved in defense against intracellular bacteria, inflammation, and some autoimmune conditions (e.g. celiac disease, psoriasis, rheumatoid arthritis, Hashimoto's thyroiditis, Graves' disease, type I diabetes, Sjögren's syndrome, Crohn's disease, multiple

sclerosis), we may expect to find these diseases as risk factors for Th1-type PTCLs.[103, 106] Th2 cells, which play a role in allergy, atopy, and defense against helminths, may be associated with diseases such as allergic dermatitis, eczema, asthma, sinusitis, inflammatory bowel disease, and other allergies.[106] This highlights the importance of conducting epidemiologic studies within specific subtypes of PTCL, as opposed to looking at the group as a whole.

*Overview of canine T zone lymphoma*

TZL, a histologic subset of PTCL-NOS, is rare in people, but comparatively common in dogs, accounting for about 12% of all canine lymphoproliferative disorders. TZL is characterized by small or intermediate size lymphocytes that expand the paracortex and medullary cords of the lymph node but do not efface the nodal architecture.[7] TZL can also be uniquely identified by the loss of the pan-leukocyte antigen CD45 on T cells using flow cytometry. In dogs, the median age at diagnosis is 10 years and the disease follows an indolent course, with average survival exceeding 2 years.[28] There is no standard treatment for TZL, but chlorambucil and/or prednisone are generally preferred over multi-agent chemotherapy due to a potential increase in overall survival.[28]

*Risk factors for canine T zone lymphoma*

No studies have assessed risk factors for TZL, but over 40% of cases occur in Golden Retrievers, supporting the genetic risk seen in human PTCL. As peripheral T cells, TZL cells are thought to be mature T cells. They also exhibit an activated phenotype, characterized by increased expression of CD21 and CD25.[6] These findings suggest a likely role of both genetics and chronic antigen stimulation in the development of TZL. By studying epidemiologic and

genetic risk factors for TZL, we can gain insights into the mechanisms underlying its pathogenesis. The Golden Retriever population provides an excellent opportunity to study these risk factors in a more genetically homogeneous population.

**Environmental and genetic risk for mature, antigen-driven lymphoproliferative disorders**

Overall, B-CLL and TZL share many characteristics, outlined in Figure 1.3. Both diseases have an indolent behavior, so they have vague clinical signs and are often diagnosed incidentally. In addition, they have good overall prognoses, with longer survival than more aggressive lymphoma subtypes. While B-CLL and TZL arise from a different cell of origin (B vs. T), they come from a similar developmental stage, both stemming from mature, peripheral lymphocytes. There is evidence of both genetic and antigen-driven components to these diseases, both from biologic and epidemiologic studies. Family history of a hematopoietic neoplasm in a first-degree relative is a risk factor for both B-CLL and PTCL.[10, 63, 103] TZL cells have high expression of markers of T cell activation and PTCL has been associated with autoimmune diseases (celiac disease, psoriasis) and atopic diseases (allergy, eczema).[6, 63, 103] Similarly, B-CLL patients exhibit restricted $IgV_H$ gene usage, and both hepatitis C virus seropositivity and farm work have been identified as risk factors.[1-4, 63, 72] These factors point to a role of chronic antigen stimulation in the pathogenesis of B-CLL and TZL. Taken together, it is likely both genetic and environmental factors play a role in the etiology of these diseases. By studying these diseases in parallel, we can increase our understanding of potential shared mechanisms underlying the development of indolent, mature, peripheral LPDs.

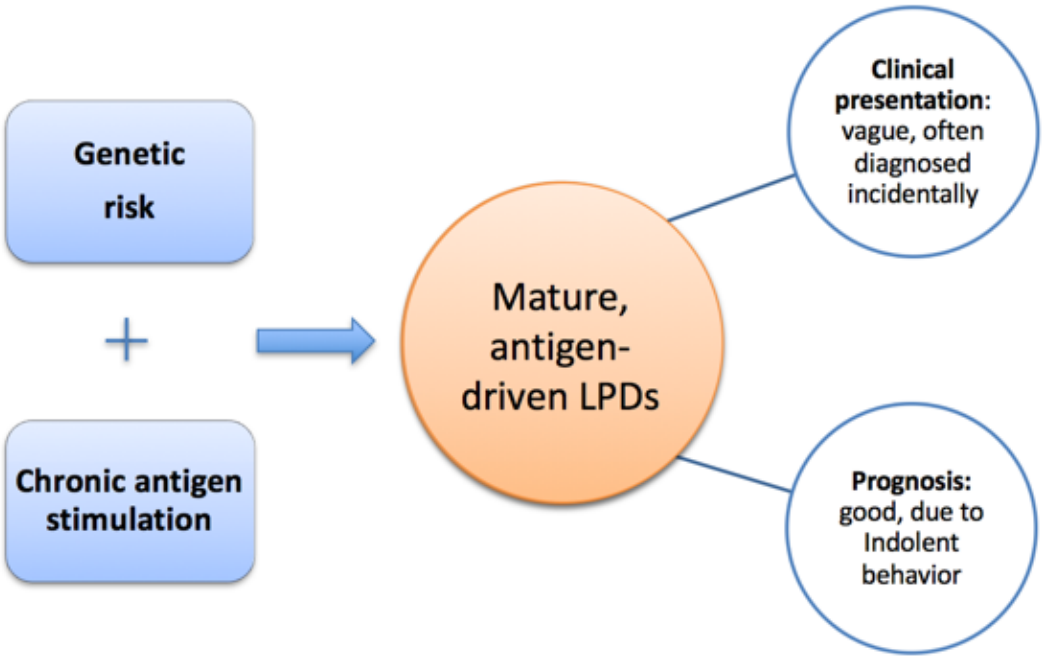



**Figure 1.3. Similarities in etiology and clinical presentation for mature, antigen-driven lymphoproliferative disorders such as B-CLL and TZL.**

# CHAPTER 2: AIM 1 – EVALUATE CANINE B CELL CHRONIC LYMPHOCYTIC LEUKEMIA (B-CLL) AS A MODEL FOR HUMAN B-CLL[1]


**Summary**

*Background*: B-cell chronic lymphocytic leukemia (B-CLL) is the most common hematopoietic malignancy among humans in the developed world and the primary risk factor is genetic. Dogs also develop B-CLL, but there is no systematic description of the disease in dogs. Understanding the epidemiology of B-CLL in dogs may help practitioners recognize the disease and position the dog as a model for future genetic studies.

*Objectives*: To describe B-CLL presentation in dogs, its clinicopathologic findings, and breed predisposition.

*Animals*: Four hundred and ninety-one dogs with B-CLL and 5,673 control dogs with suspicion of a lymphoproliferative disorder.

*Methods*: Retrospective cross-sectional study of dogs for which samples were submitted to the Colorado State University Clinical Immunology Laboratory for immunophenotyping between 2010 and 2014. To assess breed predilection, dogs with B- CLL were compared to those with suspicion of other LPDs using logistic regression.

*Results*: The median age was 11 years with no sex predilection. Half of the dogs presented with peripheral lymphadenopathy or splenomegaly and 26% had anemia. Eleven small-breed dogs had significantly increased odds of B-CLL. In addition, English Bulldogs had

an increased risk and a unique presentation: these dogs were diagnosed at a median of 6 years and expressed lower class II MHC and CD25.

*Conclusions*: B-CLL is overrepresented in small-breed dogs. Future genetic studies of these breeds may identify genetic risk factors. The unique presentation of English Bulldogs provides evidence of multiple forms of this disease. Additional studies are necessary to determine whether presenting signs are associated with survival.

**Introduction**

B-cell chronic lymphocytic leukemia (B-CLL) is the second most common leukemia affecting people in the United States.[107] This disease is well characterized and a strong genetic risk has been identified.[10] Dogs also develop B-CLL and likely have associated genetic risk factors. Because of the discrete genetic pools maintained by selective breeding, the dog may be a useful model for studying genetic risk factors for this disease. Little is known, however, about the population of dogs that develop B-CLL. Our study aims to further describe B-CLL in dogs by providing a systematic description of the typical patient presentation and clinical signs as well as exploring potential breed predisposition for the disease.

There are no consensus criteria for diagnosis of B-CLL in dogs, nor has it been determined if there is a solid tissue equivalent for this disease. Neoplasms involving canine B-cells do not express CD5, which is used to distinguish B-CLL from other B-cell lympho-proliferative disorders (LPDs) in people.[89] Prior studies used a variety of criteria for CLL diagnosis, including the use of morphologic appearance, immunophenotype, or both, with some studies excluding cases with substantial lymphadenopathy.[89, 90, 108] However,

lymphadenopathy and splenomegaly are part of the staging criteria used in humans,[109] and excluding these cases may exclude patients with advanced stage B-CLL. Because of the subjective nature of cytologic assessment of lymphocytes and the severity of lymphadenopathy, we previously carried out an analysis of B-cell leukemia characterized by small cells based solely on flow cytometric (FC) features to determine if objective criteria alone could be used to diagnose and predict prognosis in cases of B-cell leukemia.[93] Lymphadenopathy and splenomegaly were noted in 40 and 29% of cases, respectively, but these were not used as exclusion criteria. The results indicated that small cell size predicted an indolent clinical outcome in cases of B-cell leukemia and lymphoma. Although we did not specifically label this disease B-CLL, the indolent clinical course is consistent with this diagnosis.

In people, some patients experience a more progressive form of CLL, which has been associated with cell surface phenotype and $IgV_H$ mutational status.[110, 111] Prior studies have shown cell surface phenotype, including class II MHC expression, is predictive of prognosis in dogs with B-cell lymphomas.[93, 112] However, no research has directly assessed B-CLL in dogs. Because an unmutated phenotype is predictive of poor survival in people,[110] expression of surface markers indicative of B-cell maturation and activation, such as class II MHC and CD25, may have prognostic relevance in dogs with B-CLL.[113-115]

The goal of our study was to describe the clinical, immunophenotypic and breed characteristics of a large cohort of dogs with B-CLL. Our findings provide a basis on which to establish the dog model of B-CLL, and may provide clinicians and clinical pathologists with information that aids in the diagnosis of this disease. In addition, our findings can be used as a

prelude for future studies evaluating genetic risk factors as well as identifying discrete subtypes of B-CLL that may have prognostic relevance.

## Methods

*Study population*

A retrospective cross-sectional study was performed using samples submitted to the Colorado State University Clinical Immunology (CSU-CI) Laboratory. The CSU-CI laboratory (http://csu-cvmbs.colostate.edu/academics/mip/ci-lab/Pages/default.aspx) receives samples from dogs with suspicion of LPD that are submitted by reference laboratories and clinicians. For this study, all dogs were included that had samples submitted to the CSU-CI laboratory for immunophenotyping between 9-17-2010 and 6-10-2014. These samples included blood as well as aspirates of lymph nodes, bone marrow, masses and cavity fluid. If multiple samples were received for a given dog, only the first diagnostic sample was included.

*Flow cytometry*

Flow cytometry was carried out as previously described.[6] Samples acquired before May 11, 2012 were analyzed using the antibody combinations listed in panel 1 (Table 2.1) using a single laser Coulter XL (Beckman Coulter Inc., Brea, CA). Samples submitted after May 11, 2012 were analyzed with the antibody combinations listed in panel 2 (Table 2.1) using a 3-laser Coulter Gallios (Beckman Coulter Inc., Brea, CA). All data analysis was carried out using Kaluza Analysis Software (Beckman Coulter Inc., Brea, CA). Level of class II MHC expression on B cells was determined by the median fluorescence intensity (MFI) of staining on gated B cells in tube 3, panel 2 (Suppl. Figure 2.1). The CD21 and CD25 antibodies were not together in

the same staining reaction, and CD25 expression level was determined by gating on small lymphocytes in tube 2, panel 2, and excluding cells that expressed CD3, CD4, CD5 and CD8 (Suppl. Figure 2.2).

**Table 2.1. Antibody panels used for immunophenotyping.**

**Panel 1 (2 color)**

| Tube | Antibody specificity and fluorochrome |
|---|---|
| 1 | None |
| 2 | M* IgG1-FITC / CD45-PE |
| 3 | CD18-FITC / M IgG1-PE |
| 4 | CD4-FITC / CD8-PE |
| 5 | CD5-FITC / CD21-PE |
| 6 | CD3-FITC / CD45-PE |
| 7 | CD4-FITC / CD14-PE |
| 8 | Class II MHC-FITC / CD34-PE |

**Panel 2 (multicolor)**

| Tube | Antibody specificity and fluorochrome |
|---|---|
| 1 | M IgG1-FITC / M IgG1-PE / M IgG1-Alexa 647 / M IgG1-Alexa 700 / M IgG1-PE-750 / M IgG1-Pacific Blue |
| 2 | CD3-FITC / CD25-PE / CD5-APC / CD8-Alexa 700 / CD4-Pacific Blue |
| 3 | Class II MHC-FITC / CD22-PE / CD21-Alexa 647 |
| 4 | Class II MHC-FITC / CD34-PE / CD5-APC - CD14-PE-Alexa 750 |
| 5 | Class II MHC-FITC / CD18-PE / CD5-APC / CD14 PE-Alexa 750 / CD4-Pacific Blue |
| 6 | CD5-FITC / CD45-PE / CD21-Alexa 647 |

*M= mouse

Unless otherwise noted, all antibodies were purchased from AbD Serotec. Clones are as follows: CD45 = YKIX716.13, CD18 = YFC118.3 (human CD18), CD4 = YKIX302.9, CD8 = YCATE 55.9, CD5 = YKIX322.3, CD21 = CA2.1D6, CD22 = RFB4 (human CD22, purchased from AbCam), CD3 = CA17.2A12, CD14 = UCHM (human, used in panel 1) and CD14 = TUK4 (human, used in panel 2), class II MHC = YKIX334.2, CD34 = 1H6, CD25 = P2A10 (purchased from eBiosciences)

*Case definition*

The B-CLL cases were identified by flow cytometry of peripheral blood (Figure 2.1A,B). All patients were required to have a complete blood count (CBC) performed within 48 hours of sample acquisition. Only samples with >5,000 lymphocytes/μL were included. The flow cytometric criteria included homogeneous expansion (>60%) of "small" CD21+ lymphocytes. Cells were considered "small" if the ratio of the geometric mean of forward scatter of B cells to neutrophils was <0.55. The percentage of CD21+ lymphocytes was calculated as the number of CD21+ cells divided by the total number of B (CD21+) and T (CD4+ and CD8+) cells in the sample. Dogs that met this case definition were shown in a prior study to have an indolent

disease course consistent with human B-CLL.[93] No cases with expansion of CD34+ cells were included in this study. In addition to flow cytometric classification, a subset of B-CLL cases had a cytology report reviewed by a board-certified clinical pathologist (Figure 2.1C). These cases were considered "cytology-confirmed" if the report described the majority of cells as "mature", "small", having "condensed chromatin", noted "definitive or possible CLL", or some combination of these descriptors.

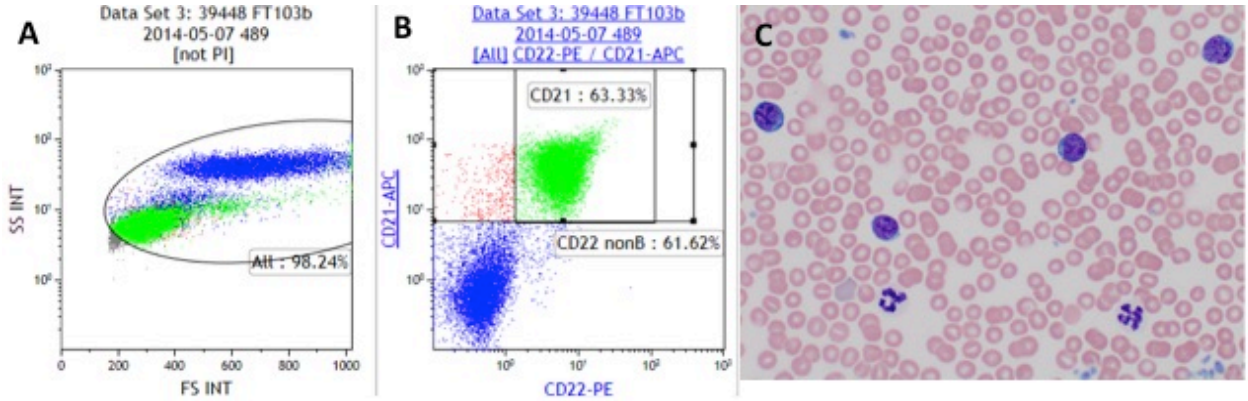



**Figure 2.1. Diagnosis of B-CLL.** Panels A and B: Flow cytometry considered diagnostic for B-CLL. The green cells indicate the neoplastic population. (A) Neoplastic cells are seen in the bottom left, indicating small size (low side and forward scatter). (B) Neoplastic cells have high expression of both CD21 and CD22, indicating the cells are B cells. (C) Blood smear considered diagnostic for B-CLL. Lymphocytes are round with condensed chromatin, no nucleoli, and a scant amount of basophilic cytoplasm. The cells are smaller than neutrophils.

*Clinical variables*

As part of the standard submission form for the CSU-CI laboratory, signalment, physical examination (PE), and laboratory findings were obtained. Signalment included breed, sex, and age at diagnosis (rounded to the nearest year). Physical examination findings included presence of peripheral lymphadenopathy, visceral lymphadenopathy, splenomegaly, hepatomegaly, or mediastinal mass. Laboratory findings included presence of hypercalcemia or hyperglobulinemia. All were categorized as present, absent, or unknown (either not evaluated

or not indicated on the form) as denoted by the clinician completing the form. Additional data abstracted from the CBC included hematocrit and absolute counts of lymphocytes, neutrophils, platelets, and reticulocytes. Because cell counts are coded into the database without specific cross-reference to the submitting laboratory's normal reference range, conservative cut points were derived from those used by the most common submitting laboratories. Neutrophil totals were categorized as neutropenic (<2,000/μL), normal (2,000 – 12,000/μL), and neutrophilic (>12,000/μL). Platelet totals were categorized as thrombocytopenic (<175,000/μL with no clumps noted), normal (175,000 – 500,000/μL), and thrombocytosis (>500,000/μL). When the platelet count was below the minimum cut-off and platelet clumps were described, these cases were excluded from analysis of platelet numbers. Hematocrit was used to define anemia (<36%). Automated reticulocyte enumeration was utilized in the majority of cases and regeneration was defined as >100,000 reticulocytes/μL.[116]

*Data analysis*

Patient presentation: For all dogs with B-CLL, continuous and interval (cell count) variables were evaluated for normality, the frequency distributions of categorical data were summarized, and descriptive statistics were calculated. Dogs diagnosed with B-CLL using cytology and flow cytometry were compared to dogs diagnosed by only flow cytometry using chi-squared or Fisher's exact tests for categorical variables and Wilcoxon rank sum test for continuous variables due to skewed distributions. Data for PE findings marked as "unknown" were excluded from analyses. As a secondary analysis, patient presentation was evaluated for all dogs with board-certified clinical pathologist-reviewed cytology reports, comparing reports

that were considered cytology-confirmed (as described above) to those for which the cytology report did not meet our criteria for a cytologic diagnosis of B-CLL. For these comparisons, Wilcoxon rank sum test and Fisher's exact test were used.

Immunophenotype: Data on CD25 and class II MHC expression were evaluated for dogs with B-CLL. These data only were available for samples received after 5-11-2012 (357 dogs) due to the change in flow cytometer. The percentage of B cells expressing CD25 was categorized into tertiles because of an underlying U-shaped distribution. Class II MHC expression was measured as the MFI on gated B cells and categorized based on the median MFI for all dogs with B-CLL.

Breed: For breed analyses, dogs reported to be a specific breed were assumed to be purebred unless otherwise noted. All "designer" breeds (e.g. "Goldendoodles") were placed in the "mixed breed" category. Only breeds with $\geq$30 total submissions during the study period were considered for analysis. All breeds with <30 submissions, or where breed was not indicated, were combined as "other." Odds ratios (ORs) and 95% confidence intervals (CIs) were calculated using a logistic regression model comparing B-CLL cases to all other immunophenotyping submissions suspected of LPDs by breed. For breeds with no B-CLL cases, Fisher's exact test was used. Separate analyses were conducted using mixed breeds and Labrador Retrievers as the reference group. The reference breeds were chosen *a priori*, with mixed breeds believed to be representative of all breeds combined, and Labrador Retrievers chosen as a well-defined and more homogeneous reference group.

From the breeds with significantly increased odds of B-CLL, the six breeds with the highest number of B-CLL cases were further examined to evaluate breed-specific patient

presentation. Because mixed breed B-CLL cases were most similar to B-CLL cases overall with respect to signalment, PE findings and CBC findings, they were used as the reference group for statistical analysis. Only variables related to signalment, PE, and laboratory findings for which ≥50% of B-CLL cases had data were evaluated. Wilcoxon rank sum test and Fisher's exact test were calculated as appropriate. All statistical analyses were conducted using SAS 9.3 (SAS Institute Inc., Cary, NC).

**Results**

*Patient presentation*

A total of 6,164 unique clinical samples were submitted to the CSU-CI laboratory for immunophenotyping by flow cytometry due to suspicion of LPD during the period 9-17-2010 through 6-10-2014. Submissions represented 48 states (ME and SD not represented), 1,046 clinics, 23 veterinary teaching hospitals, and 15 reference laboratories. Four-hundred ninety-one dogs (8%) met our case definition for B-CLL, leaving 5,673 dogs in our reference population. Board-certified pathologist-reviewed peripheral blood cytology reports were available for 333 dogs with B-CLL (68%), and 83% of these fit our criteria for cytology-confirmed B-CLL. Of the remaining 17% (55 cases), ten were described as "acute leukemia" by the reviewing pathologist. The remaining cases did not include a description of blast cells, but because cells were intermediate in size or the chromatin was described as immature, these cases did not meet our strict definition of B-CLL. Dogs with cytology reports not suggestive of B-CLL were significantly more likely to have peripheral lymphadenopathy (60% vs. 41%; p=0.04; data not shown) than dogs with cytology reports confirming B-CLL. Otherwise, there were no

significant differences between dogs with and without cytology reports confirming B-CLL with regard to patient signalment or presenting signs.

When comparing cytology-confirmed B-CLL cases to cases diagnosed exclusively by flow cytometry, no clinically meaningful significant differences were noted when evaluating breed, signalment, or presenting signs (Table 2.2). Lymphocyte total differed significantly between the two groups, but both groups had a median substantially above the normal range. In addition, all five neutropenic dogs were in the cytology-confirmed group, leading to a statistically significant difference in neutrophil count.

**Table 2.2. Signalment, physical examination findings, and laboratory findings for 491 cases of B cell chronic lymphocytic leukemia (B-CLL)**. Findings are compared between cases that are confirmed by a board-certified clinical pathologist cytology review of peripheral blood and those diagnosed by flow cytometric immunophenotyping only.

| | | | Dogs with available data (%) | Total CLL Cases (n=491) | | Cytology-confirmed cases (n=278) | | Flow cytometry only cases (n=213) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | n | % | n | % | n | % |
| **Signalment** | **Sex** | Male (intact or neutered) | 100 | 242 | 49.7 | 142 | 51.3 | 100 | 47.6 |
| | | Female (intact or neutered) | | 245 | 50.3 | 135 | 48.7 | 110 | 52.4 |
| | **Age,** *Median (IQR)* | | 98.4 | 11.0 | (9 - 13)* | 11 | (9 - 13) | 11 | (9 - 13) |
| **Veterinarian-reported Physical Exam Findings** | **Peripheral Lymphadenopathy** | | 62.7 | 141 | 45.8 | 68 | 41.2 | 73 | 51.0 |
| | **Splenomegaly** | | 46.0 | 115 | 50.9 | 63 | 52.5 | 52 | 49.1 |
| | **Hepatomegaly** | | 41.1 | 58 | 28.7 | 32 | 29.1 | 26 | 28.3 |
| | **Mediastinal Mass** | | 34.0 | 5 | 3.0 | 3 | 3.2 | 2 | 2.7 |
| | **Visceral Lymphadenopathy** | | 35.8 | 41 | 23.3 | 23 | 23.5 | 18 | 23.1 |
| **Veterinarian-reported Laboratory Findings** | **Hyperglobulinemia** | | 62.7 | 81 | 26.3 | 50 | 27.3 | 31 | 24.8 |
| | **Hypercalcemia** | | 58.9 | 13 | 4.5 | 7 | 4.1 | 6 | 5.0 |
| **CBC findings** | **Neutrophil Total^** | Neutropenia (<2,000/ul) | | 5 | 1.0 | 5 | 1.8 | 0 | 0.0 |
| | | Normal (2,000 - 12,000/ul) | | 394 | 80.6 | 223 | 80.5 | 171 | 80.7 |
| | | Neutrophilia (>12,000/ul) | 99.6 | 90 | 18.4 | 49 | 17.7 | 41 | 19.3 |
| | **Platelet Total** | Thrombocytopenia (<175,000/ul and no clumps noted) | | 28 | 6.9 | 16 | 5.8 | 12 | 5.7 |
| | | Normal (175,000-500,000/ul) | | 310 | 76.7 | 179 | 64.6 | 131 | 61.8 |
| | | Thrombocytosis (>500,000/ul) | 82.3 | 66 | 16.3 | 38 | 13.7 | 28 | 13.2 |
| | **Anemia** | Hematocrit <36 | 99.4 | 127 | 26.0 | 68 | 24.5 | 59 | 28.1 |
| | **Lymphocyte Total¥**, *Median cells x10$^3$/ul (IQR,* | | 100 | 24.6 | (14.0 - 49.4) | 27.0 | (16.1 - 51.4) | 19.9 | (12.6 - 45.8) |

^: p<0.05 comparing cytology-confirmed cases to flow cytometry cases using Fisher's exact test
¥: p<0.05 comparing cytology-confirmed cases to flow cytometry cases using Wilcoxon rank sum test
*: values for age and lymphocyte total are given as median (IQR)

Table 2.2 shows signalment, PE, and CBC findings for all 491 B-CLL cases in the study. Overall, there was a 1:1 ratio of males to females, with 242 males (24 intact, 212 neutered, 6

unknown) and 245 females (3 intact, 239 neutered, 3 unknown). The median age at diagnosis was 11 years (IQR, 9-13; range, 2-17). Data on peripheral lymphadenopathy were available for 63% of dogs; of these, 141 (46%) had peripheral lymphadenopathy. Of the dogs for which PE findings were reported, 51% had splenomegaly, 29% had hepatomegaly, 3% had a mediastinal mass, and 23% had visceral lymphadenopathy. Veterinarian-reported biochemical data were available for approximately 60% of dogs. Twenty-six percent of dogs had hyperglobulinemia and 5% had hypercalcemia.

The B-CLL case definition required a minimum lymphocyte count of 5,000/μL and >60% B-cells. Among dogs with B-CLL, the median lymphocyte count was 24,600/μL (IQR, 14,000–49,370; range, 5,000–812,544) and the median percentage of lymphocytes that were B cells was 94% (IQR, 89%-97%). Overall, neutropenia and thrombocytopenia were rare (1 and 7% of cases, respectively), but anemia was found in 26% of dogs. When present, thrombocytopenia generally was mild (median, 137,000/μL; IQR, 89,000–161,000/μL). Among the 128 anemic dogs, hematocrit ranged from 11-35% (median, 30%; IQR, 26-34%; Suppl. Table 2.1). Reticulocyte counts were available for half of the anemic dogs (67/128); of these, 31% (21/67) had regenerative anemia (>100,000/μL reticulocytes). Most anemic dogs were not neutropenic or thrombocytopenic (98% and 85%, respectively).

*Breed*

We compared the proportion of breeds in the B-CLL population to the proportion in the population of all dogs with suspected LPD using logistic regression. Thirty-seven breeds had ≥30 submissions for suspected LPD (Figure 2.2; Suppl. Table 2.2). Together, these breeds

represented >80% of both total submissions (5108/6164) and B-CLL cases (401/491). Of the breeds with ≥30 submissions, 12 had significantly increased odds of B-CLL compared to mixed breeds. This analysis included 10 small breeds (Bichon Frise, Boston Terrier, Cairn Terrier, Cocker Spaniel, Dachshund, Jack Russell Terrier, Maltese, Pomeranian, Shih Tzu, and Yorkshire Terrier) and 2 larger breeds (English Bulldog and Pit Bull). All breeds that were significantly different from mixed breeds remained significant when compared to Labrador Retrievers (Figure 2.2; Suppl. Table 2.2); 4 additional breeds were significantly different from Labrador Retrievers (Boxer, Chihuahua, West Highland White Terrier, and Doberman).

Six large breeds had significantly decreased odds of B-CLL compared to mixed breeds: Bernese Mountain Dog, German Shepherd, Golden Retriever, Labrador Retriever, Rottweiler and Standard Poodle. Welsh Corgis also were at decreased risk (Figure 2.2; Suppl. Table 2.2). Notably, although Golden Retrievers comprised the majority of purebred submissions, only 3 cases of B-CLL were identified.

Breed-specific patient presentation for the top six breeds with statistically significant increased odds of B-CLL was compared to mixed breeds. These breeds were Shih Tzu (n=28), Pit Bull (n=15), Cocker Spaniel (n=14), English Bulldog (n=23), Jack Russell Terrier (n=14), and Dachshund (n=14). No significant differences were noted for most variables (sex, lymphocyte count, neutropenia, thrombocytopenia, and anemia; data not shown). We did not evaluate the frequency of splenomegaly, hepatomegaly, mediastinal mass, and hypercalcemia because data were available for <50% of B-CLL cases. English Bulldogs, however, presented at a significantly younger age (median 6 vs. 11 years for mixed breeds; $p < 0.001$; Figure 2.3). In contrast, Cocker Spaniels were significantly older when diagnosed (median, 13.5 years; $p=0.012$) and were

significantly less likely to present with peripheral lymphadenopathy (11% vs. 53% for mixed breeds; p=0.03; data not shown). Half of the English Bulldog B-CLL cases had data available on hyperglobulinemia. Of those, 55% were hyperglobulinemic (vs. 23% among mixed breeds; p=0.06).

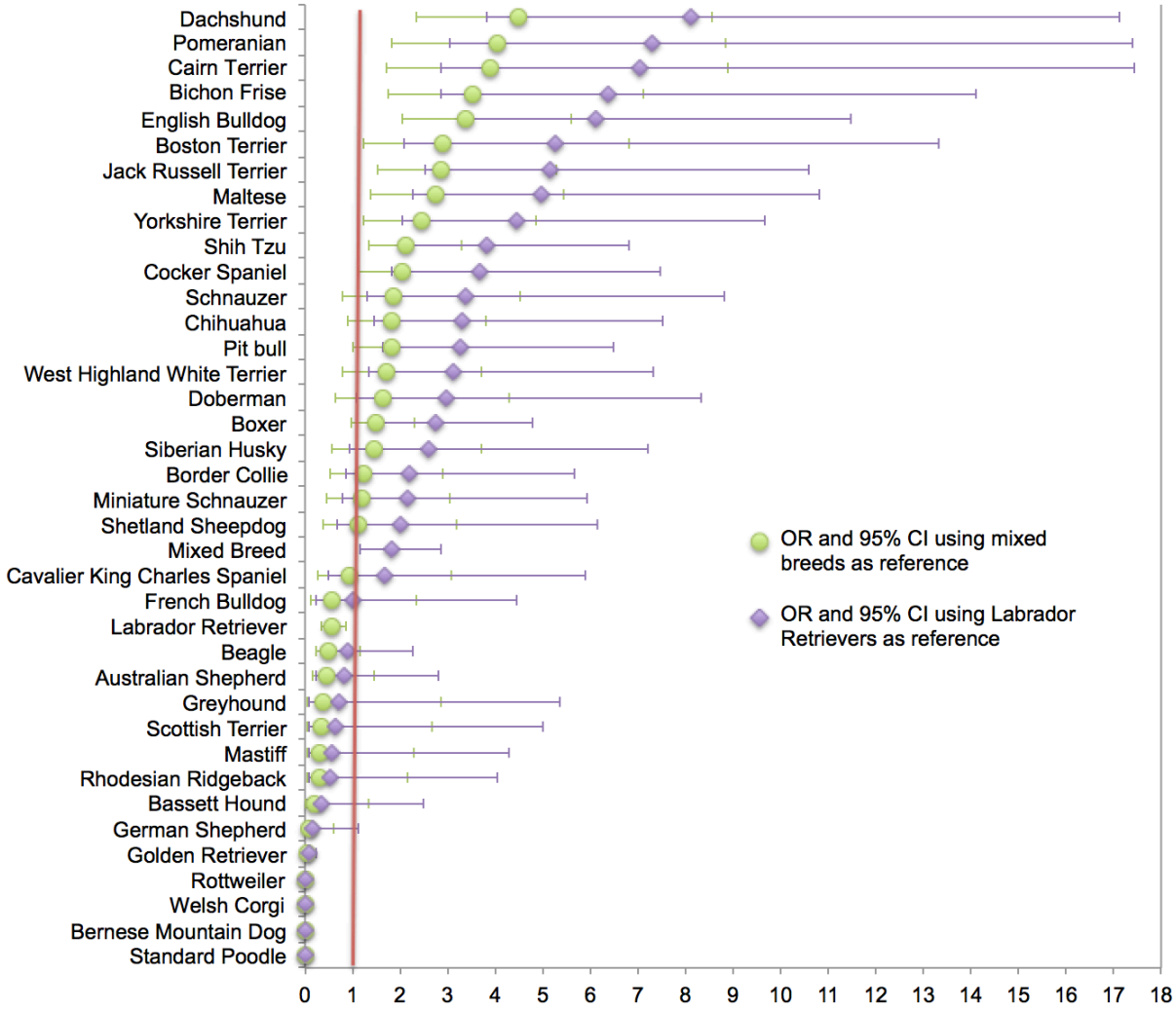



**Figure 2.2. Odds ratios and 95% confidence intervals comparing B-CLL among each breed to all other lymphoproliferative disorders, using mixed breeds and Labrador Retrievers as a reference.** Only breeds with at least 30 overall submissions were considered (37 breeds). Twelve breeds had significantly increased odds of B-CLL when compared to both mixed breeds and Labrador Retrievers: Dachshund, Pomeranian, Cairn Terrier, Bichon Frise, English Bulldog, Boston Terrier, Jack Russell Terrier, Maltese, Yorkshire Terrier, Shih Tzu, Cocker Spaniel, and Pit Bull. In addition, six breeds had significantly decreased odds of B-CLL compared to both mixed breeds and Labrador Retrievers: German Shepherd, Golden Retriever, Rottweiler, Welsh Corgi, Bernese Mountain Dog, and Standard Poodle.

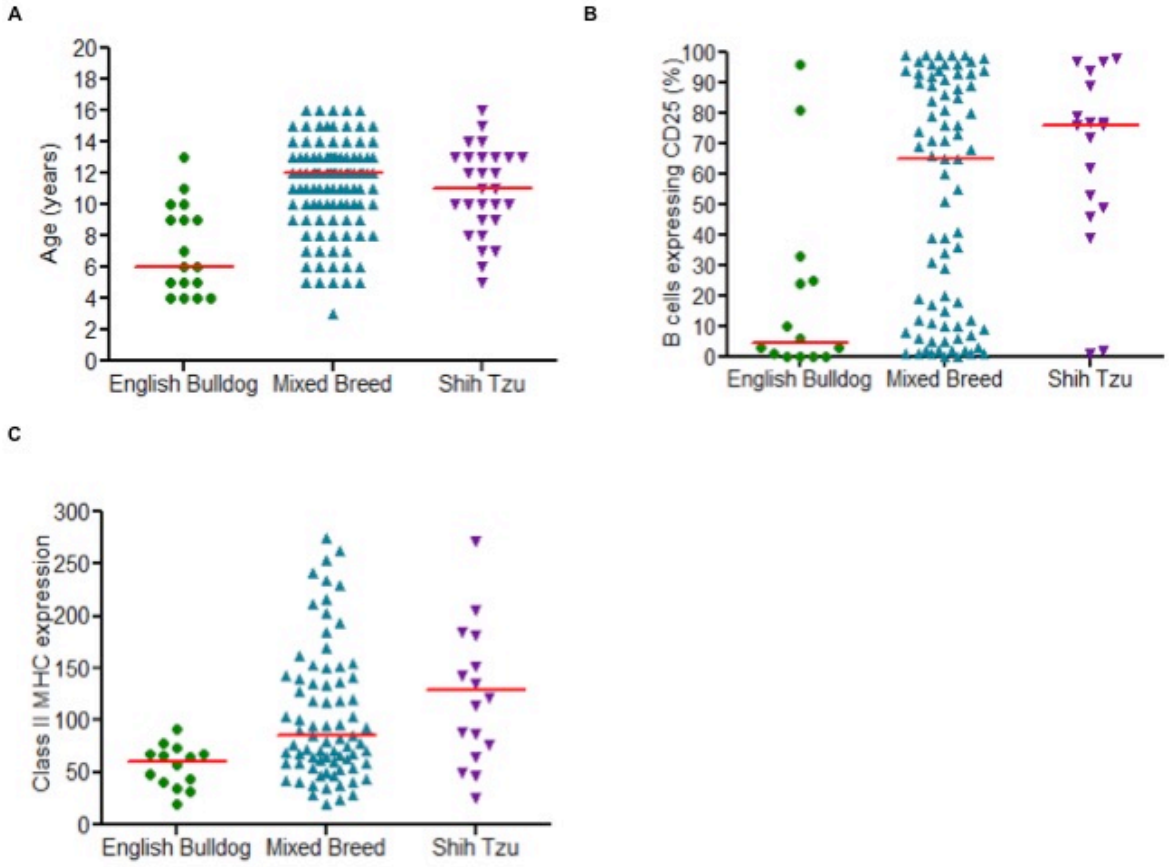



**Figure 2.3. Age (A), percent of CD25-positive B cells (B), and median class II MHC expression (C) were compared for the top six breeds with significantly increased odds of B-CLL (Shih Tzu, Pit Bull, Cocker Spaniel, English Bulldog, Jack Russell Terrier, Dachshund) vs. mixed breed dogs.** Significant differences were noted for English Bulldogs and Shih Tzus with B-CLL, and are displayed in the figure, with red horizontal lines indicating median values. Compared to mixed breed dogs, English Bulldogs presented at a significantly younger age, with lower percent of CD25-positive B cells, and decreased class II MHC expression. Shih Tzus had no significant difference in age or class II MHC expression, but were significantly less likely to have a low percentage of CD25-positive B cells.

*Flow cytometric characterization*

Among cases, we examined the expression of 2 antigens, CD25 and class II MHC, on canine B-CLL cells, comparing the top six breeds (as described above) to mixed breed dogs using Fisher's exact test. Significant differences were only noted for English Bulldogs and Shih

Tzus. The level of CD25 expression was measured as the percentage of B cells expressing CD25 (Suppl. Figure 2.2). Overall, the median percentage of CD25-positive B cells was 58% (IQR, 13%-88%; data not shown). Tertile cutpoints were used to categorize the percentage of CD25-positive B cells into low (<29% of B cells CD25-positive), medium (29%-80%), and high (>80%) groups. Compared to mixed breeds, English Bulldogs were more likely to fall into the lowest tertile (have a low percentage of CD25-positive B cells), with a median of 4.5% CD25-positive B cells ($p=0.012$; Figure 2.3). In contrast, Shih Tzus had a median of 76% CD25-positive B cells, making them less likely to fall into the low category ($p=0.023$). All cases expressed class II MHC at varying levels. English Bulldogs were significantly more likely to have class II MHC expression below the median (94% vs. 51%; $p<0.001$; median, 54.5%; Figure 2.3, Suppl. Figure 2.1). More Shih Tzus had class II MHC expression above the median (72%), but this difference was not statistically significant ($p=0.116$).

**Discussion**

Overall, 8% of samples submitted for immunophenotyping because of suspicion of LPD were diagnosed as B-CLL. Our study demonstrated that B-CLL in dogs shows strong breed-specific risk relative to the population of dogs with suspected LPD. Ours is the first published study to evaluate this association, and represents the largest population of dogs with B-CLL in the veterinary literature. Several breeds with an increased prevalence of other LPDs (Golden Retriever, German Shepherd, Rottweiler) are rarely diagnosed with B-CLL (Figure 2.2). With the exception of English Bulldogs and Pit Bulls, all of the breeds with significantly increased odds of B-CLL were small breed dogs. This finding could be due to a shared genetic risk in

small breed dogs[117, 118] or differences in the average lifespan of small vs. large breed dogs. Small breed dogs tend to live longer, allowing them a prolonged opportunity to develop clinical disease. In contrast, many larger breed dogs are nearing the end of their lifespan around the average age of B-CLL diagnosis and the vague clinical signs of B-CLL could be misinterpreted as signs of old age or could be masked by more severe health problems. One resistant breed, however, the Golden Retriever, is highly represented among T-zone lymphoma, which also is a disease that affects older dogs (median age, 10.5 years).[6]

By evaluating breed-specific presentation, we determined that certain breeds have a unique presentation of B-CLL (Figure 2.3). Compared to mixed breed dogs, English Bulldogs presented at a significantly younger age, and with decreased class II MHC and CD25 expression. Prior research has shown that CD25 is expressed on mature or memory B cells,[114, 115] Thus it is plausible B-CLL in English Bulldogs arises from a naïve, unactivated B cell. The low level of class II MHC expression in English Bulldogs also supports this notion, and suggests that further investigation of the activation state of these cells is warranted. Thus, our results suggest that English Bulldogs exhibit a unique B-CLL phenotype. Additional study of the B cells in this breed, including $IgV_H$ mutation status, is underway.

We found that B-CLL generally is a disease of older dogs that affects males and females equally (Table 2.2). Where reported, approximately half of the dogs had peripheral lymphadenopathy or splenomegaly and a quarter had hepatomegaly, visceral lymphadenopathy, or hyperglobulinemia. Hypercalcemia and mediastinal mass were rare, suggesting that it would be prudent to seek an alternative cause for these clinical signs if they are present in a dog with B-CLL. Serum protein electrophoresis data were not available in our

population, but a previous study found 68% of dogs with hyperglobulinemia had a monoclonal gammopathy.[88] With the exception of lymphocytosis, the majority of dogs had normal CBC findings. Approximately 26% of dogs were anemic on presentation, and the majority (68%) were classified as having non-regenerative anemia (Table 2.2). However, only 15% were thrombocytopenic and 1% were neutropenic. Of eight dogs with bone marrow aspirates available, all but 1 had increased small, mature lymphocytes, with infiltration ranging from 28%-65% of nucleated cells. However, only 2 of these dogs had associated cytopenias (anemia and thrombocytopenia). In humans with B-CLL, anemia attributed to autoimmune destruction is found in 3-37% of cases.[119] Because the majority of anemic dogs in our study had non-regenerative anemia with no comments about spherocytes, we suspect most had anemia of chronic disease. Of the 21 dogs with regenerative anemia, 10 had clinical pathologist-reviewed cytology reports. Only 1 of these reports mentioned spherocytes; all reviews either did not mention saline agglutination tests or mentioned that saline agglutination tests were negative (n=3). Thus, only one of these 10 dogs had suspicion for immune-mediated anemia.

The typical patient signalment seen in this study parallels that of previous studies.[88, 89, 108, 120] However, as previously mentioned, our inclusion and exclusion criteria differed from prior studies. For example, diagnostic criteria for two studies[90, 108] included morphologic appearance of cells and excluded cases with moderate to severe lymphadenopathy, which may have skewed findings toward less advanced cases. In contrast, another study[89] required cases to have a lymphocytosis of >50,000 cells/μL, which may have skewed findings toward advanced-stage cases. Importantly, many prior studies had small sample sizes and included both B- and T-CLL, making it difficult to make accurate comparisons with our cases.

One limitation of our study is that, in cases with nodal involvement, we were unable to distinguish B-CLL from other forms of B cell LPDs that include a circulating component of small, mature B cells. Such a distinction is not possible without histology and immunohistochemistry, which were unavailable for our study participants due to the retrospective nature of our study. It is therefore possible our dataset includes mantle cell lymphoma, nodal marginal zone lymphoma, splenic marginal zone lymphoma, or follicular lymphoma with leukemic involvement.[79] However, these histologic subtypes appear to be rare in dogs,[7, 8, 121] and we suspect their influence is minimal and that the majority of our cases truly are B-CLL. In addition, lymphocytosis was not found in any of the 15 cases of marginal zone lymphomas confirmed by H&E staining and IHC in another study.[28] Taken together, we believe it is likely that most of these cases were B-CLL. In ongoing studies, we are recommending lymph node biopsies from these dogs to further explore the extent to which other histologic subtypes may be represented.

Although 17% (n=55) of dogs with cytology reports did not meet our strict definition of cytology-confirmed B-CLL, the majority (71%) were consistent with B-CLL. Only 10 cases had cytology reports that contradicted a B-CLL diagnosis. When we compared patient presentation between dogs with cytology reports confirming B-CLL and the 17% that were not confirmatory, the only relevant difference noted was a higher frequency of peripheral lymphadenopathy among dogs without a confirmed diagnosis. Therefore, we do not expect this finding to substantially bias the patient presentation findings we reported, with the exception of potential overrepresentation of the frequency of peripheral lymphadenopathy.

It currently is unclear whether flow cytometric or cytologic evaluation is the more accurate method for diagnosing this disease and predicting outcome. In this study, we found no substantial differences in clinical presentation between cytology-confirmed B-CLL cases and those diagnosed with flow cytometry only (Table 2.2). Because cytologic interpretation can differ among clinical pathologists, we favor moving toward more objective criteria, such as flow cytometry, in the classification of B-CLL and other forms of leukemia.

A second limitation of the study is that the reference population for breed calculations is drawn from dogs with a suspicion of LPD. A breed predilection for lymphomas has been reported.[74, 122] Therefore, the breed distribution in our study population is likely skewed toward these breeds and may not be representative of the overall breed distribution of dogs in the United States. We chose mixed breed dogs as our primary comparison group for breed calculations because we believe they represent the genetic diversity of all dogs. Furthermore, they closely represented the overall patient presentation seen in our B-CLL population (Suppl. Table 2.3). However, they are not an ideal reference group because they are an unknown group of dogs that may vary considerably across populations. To have a better-defined comparison group, we also used Labrador Retrievers, a more homogeneous group that is more likely to be comparable across populations. However, Labrador Retrievers represented a smaller population of dogs with B-CLL (n=23) and had a lower frequency of B-CLL than the overall population (5% vs. 8%). Therefore, the associated ORs for these calculations were skewed toward larger values than when comparing to mixed breed dogs. The overall trend was the same for both comparisons, increasing our confidence that B-CLL occurs more commonly in small breed dogs (Figure 2.2).

Taken together, the results from this large population of dogs suggest that B-CLL in dogs shares many features with B-CLL in humans. Dogs are gaining recognition as models for naturally-occurring human cancers, and thereby may be useful for studying risk factors and treatments for B-CLL in humans. Furthermore, the strong breed-specific risk identified in this study poses an opportunity to study genetic risk factors for B-CLL and to identify unique subpopulations of B-CLL.

# CHAPTER 3: AIM 2 – IDENTIFY GENES ASSOCIATED WITH T ZONE LYMPHOMA


**Summary**

*Introduction:* T zone lymphoma (TZL), a histologic variant of peripheral T cell lymphoma not otherwise specified, represents about 12% of all canine lymphomas. In dogs, this disease follows an indolent clinical course with average survival of >2 years regardless of treatment, compared to <1 year with most other lymphoma subtypes. TZL appears to have a striking predilection for Golden Retrievers, which represent over 40% of cases. This finding suggests a genetic risk factor for this disease and caused us to pursue a study to identify potential genes of interest.

*Materials & Methods:* Case-subclinical-control status of privately-owned U.S. Golden Retrievers was confirmed using flow cytometry; cases were diagnosed using previously established flow cytometric criteria. Dogs were considered subclinical if they had a small population of cells phenotypically similar to TZL, but had no lymphadenopathy or lymphocytosis. Controls had no evidence of TZL. A total of 95 TZL cases, 142 subclinicals, and 101 controls were genotyped using the Illumina CanineHD BeadChip. After standard quality control and relatedness filtering, the remaining samples were analyzed using a mixed linear model adjusting for remaining population stratification using GCTA software. Significance was determined based on 95% confidence intervals, resulting in a threshold of $10^{-4}$. Associated regions were subsequently sequenced using a custom sequence capture array (NimbleGen SeqCap EZ Developer Kit) on an Illumina NextSeq 500.

*Results:* We found genome-wide significance in regions on chromosomes 8 and 14. The chromosome 14 peak included four single nucleotide polymorphisms (SNPs), with odds ratios ranging from 1.18–1.19 ($p=2.3x10^{-5}$–$5.1x10^{-5}$). This region contains multiple genes encoding hyaluronidases, including SPAM1, HYAL4, and HYALP1. Targeted resequencing of this region identified missense mutations in all three genes; the variant in SPAM1 was predicted to be damaging. These mutations were also identified in a study of mast cell tumors in Golden Retrievers. The chromosome 8 peak contained seven SNPS significantly associated with TZL, with odds ratios ranging from 1.24–1.42 ($p=2.7x10^{-7}$–$7.5x10^{-5}$). This region includes genes encoding thyroxine deiodinase and thyroid stimulating hormone receptor, both of which are involved in thyroid hormone regulation. It also contains genes involved in vesicle trafficking and degradation of misfolded proteins. No coding mutations were found with targeted resequencing, but hundreds of variants were identified in the region, and may play a regulatory role for all or some of the genes.

*Conclusions:* The pathogenesis of canine TZL may be related to hyaluronan breakdown. Hyaluronan interacts with CD44, a cell surface glycoprotein which is expressed on both T cells and mast cells. This binding may cause increased hyaluronan turnover, which leads to the production of pro-inflammatory and pro-oncogenic by-products. In addition, hyaluronan can directly promote inflammation by activating the inflammasome. The association on chromosome 8 may indicate thyroid hormone, accumulation of misfolded protein, and/or changes in antigen presentation could influence development of TZL. Future work is needed to elucidate these mechanisms.

**Introduction**

Peripheral T cell lymphomas (PTCLs) account for less than 10% of human non-Hodgkin's lymphomas worldwide.[20, 21, 123] This is a heterogeneous group of malignancies which generally exhibit poor prognosis and resistance to standard chemotherapy. Overall, 5-year survival is only 32%.[97, 98, 100, 101] Gene expression profiling has enabled the subtyping of PTCL based on cell of origin, allowing the identification of molecular diagnostic and prognostic signatures.[105, 124] However, the etiology of PTCLs is unclear, largely due to the low incidence within subtypes, which limits our ability to gather adequate numbers of cases.

Among dogs, PTCL accounts for 30% of lymphomas, enabling us to enroll larger sample sizes of individual subtypes.[40] For example, T zone lymphoma (TZL) is a rare subtype of human PTCL, making it difficult to study. However, TZL is relatively common in dogs, representing about 12% of all canine lymphomas.[40, 125] In dogs, this disease follows an indolent course with average survival of >2 years regardless of treatment, compared to <1 year with most other lymphoma subtypes.[6, 8, 28] Few studies have investigated the pathogenesis of canine TZL. However, it has been noted that TZL appears to have a striking predilection for Golden Retrievers, with over 40% of cases occurring in this breed.[6] This finding suggests a genetic risk factor for this disease, and caused us to pursue a study to identify potential pathways of interest. From an evolutionary perspective, the development of dog breeds was relatively recent. Because of this, there is reduced genetic variation within breeds of dogs,[37, 38] allowing us to use smaller sample sizes and fewer genetic markers when evaluating genetic risk factors for canine diseases.[38, 42]

The objective of this study was to identify genetic risk factors for canine TZL using a genome-wide association study (GWAS) and subsequent targeted sequencing. Results from this study will provide insight into the etiology and underlying risk for developing this disease.

## Methods

*Study participants*

All dogs were recruited from the privately owned pet population in the United States. The study was conducted with approval from the Colorado State University (CSU) Institutional Animal Care and Use committee. Peripheral blood samples were submitted to the CSU Clinical Immunology laboratory from October 2013 through May 2015. Flow cytometric analysis of peripheral blood samples was used to diagnose dogs as cases, subclinicals, or controls. Flow cytometry was carried out as previously described[6] and samples were analyzed with the antibody combinations listed in Suppl. Table 3.1 using a 3-laser Coulter Gallios (Beckman Coulter Inc., Brea, CA). All data analyses were carried out using Kaluza Analysis Software (Beckman Coulter Inc., Brea, CA). *Case* definition included a homogeneous expansion of $CD5^{+}CD45^{-}$ T cells as well as the presence of lymphocytosis (defined as >5,000 lymphocytes/μL) and/or lymphadenopathy (Suppl. Figure 3.1A). *Controls* were at least 9 years of age, had no history or suspicion of a lymphoproliferative disease (including no lymphadenopathy or lymphocytosis), and had no population of $CD5^{+}CD45^{-}$ T cells identified by flow cytometry (Suppl. Figure 3.1C). An additional group of dogs, *subclinicals*, were identified among dogs submitted as controls. This group also had no history or suspicion of a lymphoproliferative disease, but had a small population of $CD5^{+}CD45^{-}$ T cells on flow cytometry (>1% of total T

cells) without the presence of a lymphocytosis or lymphadenopathy (Suppl. Figure 3.1B). We have observed that over 30% of "normal" aged Golden Retrievers are subclinicals (Hughes et al., under review).

*Genome-wide association mapping*

Genomic DNA was extracted from white blood cell pellets of peripheral blood samples using the QIAamp DNA blood Midi Kit (QIAGEN Inc., Germantown, MD). DNA quality and quantity were assessed using NanoDrop (Thermo Scientific, Wilmington, DE) and only samples with 1) a concentration of at least 35 ng/μL, 2) over 1,000ng of DNA total, and 3) an A260/280 ratio of 1.7-2.0 were submitted for genotyping. Genotyping was performed at GeneSeek Inc. (Lincoln, NE) using the Illumina 170K CanineHD BeadChip.[126] PLINK software[45, 127] was used to perform data quality control, removing individuals with call rates <97.5% and single nucleotide polymorphisms (SNPs) with call rates <97.5% or minor allele frequency (MAF) < 5%. Only autosomal chromosomes were analyzed.

Multidimensional scaling plots were assessed using PLINK to ensure there were no distinct groupings based on phenotype (case/subclinical/control), which would indicate population stratification and/or residual confounding. While we saw no obvious deviations on these plots, we were concerned for bias based on European vs. American descent due to apparent divergence in this breed.[42, 128] To determine whether any of our dogs were likely of European descent, we downloaded a publicly available dataset[128] including both European and American Golden Retrievers, conducted the same quality control protocol as described above, and merged the two datasets. We then created multidimensional scaling plots to determine

which dogs in our dataset clustered with the known European dogs. Dogs with a value for the first cluster <2.5 standard deviations from the mean value for known European dogs were removed (Suppl. Figure 3.2).

Genome-wide complex trait analysis (GCTA) software[46] was used to estimate a genetic relationship matrix (grm) and remove highly related individuals (one dog was removed for each pair of dogs with the same phenotype and a grm value of 0.25 [half-sibling level]). The disease-genotype association was estimated using GCTA, adjusting for the first principal component of the grm in a mixed linear model to correct for cryptic relatedness.[129] Quantile-quantile (QQ) plots with 95% confidence intervals (CIs) calculated based on the beta distribution of observed p-values were created to assess possible genomic inflation and to establish suggestive significance levels.

It is currently unclear whether subclinicals share some or all of the genetic predisposition for TZL with cases, or they simply represent normal variation among controls. To determine in which category they are most genetically informative, we performed separate association studies comparing 1) combined cases and subclinicals vs. controls, 2) cases vs. combined subclinicals and controls, and 3) pairwise comparisons (case vs. control, case vs. subclinical, and subclinical vs. control). QQ plots were used to determine which analyses had enough power to be evaluated. After peaks of interest were identified, we used GCTA to conduct a conditional GWAS. By adjusting for the genotype of the top SNP of each peak of interest, we can evaluate whether the other significantly associated SNPs in that peak are statistically independent from the top SNP (i.e. whether there are one or multiple peaks).

*Haplotype block definition and association analysis*

Haplotype blocks for associated loci were defined based on boundaries identified both by clumping analysis in PLINK and $R^2$-based linkage disequilibrium (LD) analysis in Haploview.[130] For clumping analysis, the dataset was subsetted to a region including all SNPs with $R^2 \geq 0.2$ from the top SNP on that chromosome. Because of the genomic structure of dogs, a maximum window size was set to 5000 kb. For each block, their haplotype, frequencies, chi-square test, and p-value were obtained using PLINK. We additionally created a risk haplotype score to assess cumulative risk; dogs were scored from zero to four based on the number of risk haplotypes present for the associated regions.

*Targeted sequencing*

Sixteen dogs (10 cases, 3 subclinical, 3 controls), selected for optimal haplotype representation (i.e. to represent risk – risk, risk – non-risk, and non-risk – non-risk for each haplotype) and distribution in multidimensional scaling plot, were sequenced across the associated genomic regions (Suppl. Table 3.2). A custom sequence capture array was designed (NimbleGen SeqCap EZ Developer Kit; Roche Diagnostics Corporation, Indianapolis, IN) to cover the top associated regions (16.1 Mb total; CanFam 3.1 cfa8:51,700,000-54,800,000, cfa14:8,000,000-16,100,000, cfa29:7,600,000-12,500,000). Regions were chosen to include all SNPs with $R^2 \geq 0.2$ from the top SNP. Standard indexed Illumina libraries were prepared with the KAPA HyperPlus library preparation kit (KapaBiosystems, Wilmington, MA). Targeted pooled (4 samples) libraries were captured by hybridization in solution using the custom probe pool. Estimated coverage of the 16.1Mb target region was 95%. Library constructions, pooling and

captures were performed following the SeqCap EZ HyperCap Workflow User's Guide (V 1.0) (Roche Diagnostics Corporation, Indianapolis, IN). Per suggestion from NimbleGen, developer's reagent (06684335001) was used in place of COT-1. Index-specific hybridization enhancing oligonucleotides were used. Sequencing was carried out on an Illumina NextSeq 500.

Sequencing data were pre-processed and aligned to the CanFam3.1 reference genome using FastQC[131], Samtools,[132, 133] Picard Tools,[134] Genome Analysis Toolkit,[135] and BWA-MEM,[136] as specified by Genome Analysis Toolkit best practices. Data was visualized using Integrative Genomics Viewer.[137] Genome Analysis Toolkit was used for base quality score recalibration (BaseRecalibrator), variant calling (HaplotypeCaller), and variant prioritization (VariantFiltration). Variants were additionally filtered based on adherence to the risk or non-risk haplotype (at least 80% adherence) and variants that passed this filter were annotated using SnpEff.[138] Coding SNPs were evaluated for predicted effect using PolyPhen-2.[12] Ensembl was used to perform a LiftOver to the human genome to determine whether non-coding SNPs were in potential regulatory elements.

**Results**

The target population included 95 TZL cases, 101 TZL-free control dogs >9 years old, and 142 subclinicals >9 years old (dogs with no clinical signs of TZL, but >1% of T cells were $CD5^{+}CD45^{-}$). Sixteen dogs were removed due to low genotyping rate (<97.5%; 7 cases, 5 subclinicals, 4 controls) and 6 were removed due to suspected European origin (2 subclinicals, 4 controls). After quality filtering a final dataset of 267 dogs (79 cases, 108 subclinicals, 80 controls), and 110,405 SNPs was used for association analyses.

*Subclinicals and controls indistinguishable by GWAS*

When the combined case and subclinical group was compared to controls, no p-values were outside the 95% confidence interval threshold on the QQ plot (Suppl. Fig 3.3A). In contrast, when cases were compared to the combined subclinical and control group, a group of SNPs significantly deviated from the expected distribution (Figure 3.1). Supporting this, pairwise GWAS of cases vs. controls and cases vs. subclinicals had suggestive associations for this group of SNPs, despite none of the p-values falling outside the 95% CI on the QQ plot (Suppl. Figs. 3.3B and C). This implies subclinicals and controls are similar, and the enhanced power from combining them as a reference group allows those SNPs to reach genome-wide significance. In contrast, the subclinical vs. control comparison did not share any suggestive SNPs with the case vs. control comparison, as would be expected if cases and subclinicals were similar. We thus chose to combine subclinicals and controls for our main analysis, and will reference it as the "case vs. all" comparison for the remainder of the paper.

*Top peak is near thyroid stimulating hormone receptor locus*

The strongest GWAS peak contained seven SNPs on chromosome 8 from 52,650,576–53,818,371 bp (Table 3.1). The associated allele for these SNPs was present in about 16% of cases (range 15%-25%) compared to 6% of the reference group (range 4%-12%). The top SNP (BICF2P948919; OR = 1.39, p-value = $2.66 \times 10^{-7}$) was located at 53,818,371 bp and was in strong LD ($R^2 > 0.7$) with three significantly associated SNPs in that region and moderate LD (0.25-0.6) with the other three significantly associated SNPs (Figure 3.2). Using the PLINK clumping analysis, we determined that the four SNPs in strong LD (including the top SNP) formed one

haplotype block, and the remaining three SNPs were not in strong enough LD with any other SNPs to form blocks. The p-values for all seven associated SNPs on chromosome 8 were non-significant (range 0.17-0.99) in the conditional analysis, suggesting they represent one signal (Table 3.1). The haplotype block containing the top SNP is within the non-coding region of Suppressor of Lin-12-Like Protein 1 (SEL1L) gene (Figure 3.2). Having at least one risk haplotype was substantially more common among cases (29%) vs. subclinicals or controls (12% and 7.5%, respectively).

*Targeted resequencing of the chromosome 8 region identifies potential regulatory variants*

Sequence capture of the 3Mb region on chromosome 8 in 16 dogs selected for variation in risk and non-risk haplotypes identified 814 single nucleotide variants (SNVs) and 229 insertions and deletions (indels) that passed our filters. Median coverage across the region was 131x. Three synonymous coding variants were found in the SEL1L gene (cfa8:53,771,782, cfa8:52,779,502, cfa8:53,797,623). All other identified variants were potential modifiers, including 3' UTR variants (three SNVs and one indel near CEP128, two SNVs near GTF2A1), up- and downstream gene variants, intron variants, and non-coding transcript exon variants (Suppl. Tables 3.3 and 3.4). Evaluation of the corresponding positions in the human genome determined multiple variants were in potential regulatory elements (of 685 that were converted [541 SNV, 144 indels]; based on H3K27AC marks and GeneHancer scoring). Two sets of variants were in enhancers for DIO2 (Type II Iodothyronine Deiodinase) and seven for combinations of CEP128 (Centrosomal Protein 128), GTF2A1 (General Transcription Factor IIA Subunit 1), STON2 (Stonin2), and SEL1L (Suppl. Table 3.5).

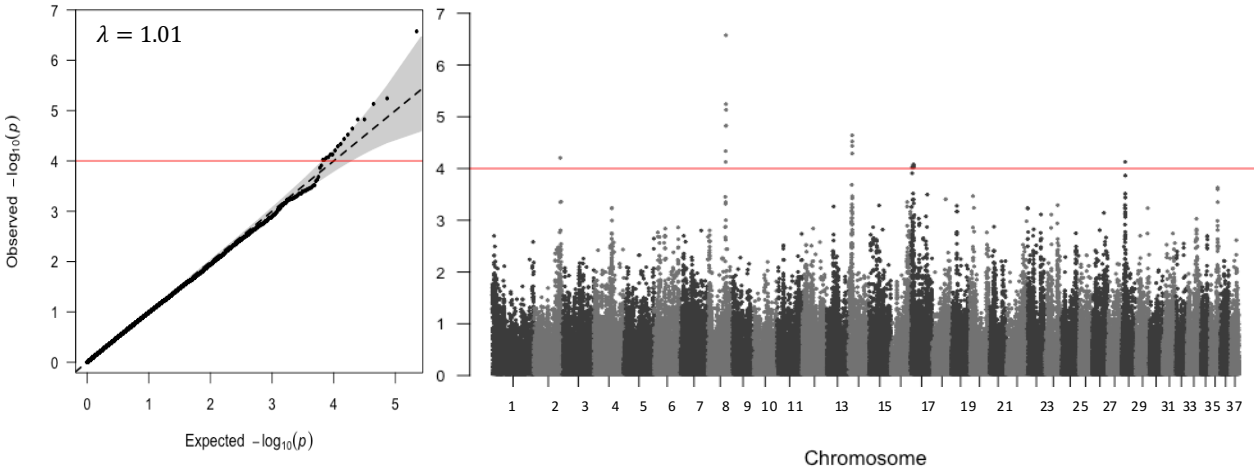



**Figure 3.1. GWAS results for TZL cases vs. combined reference (subclinicals + controls).** Left, QQ plot demonstrating observed p-values deviate from the expected at a significance level of $p<10^{-4}$. Shaded area indicates 95% confidence interval. Right, Manhattan plot showing peaks that are significantly associated with TZL at a genome-wide level of $p<10^{-4}$. The main association signals come from chromosomes 8 and 14, with less significantly associated peaks on chromosomes 2, 17, and 29.

**Table 3.1. SNPs significantly associated with TZL at the genome-wide level.** The top SNPs (smallest p-value) for the chromosome 8 and 14 peaks are bolded.

| | | | | Associated Allele Freq | | | | | Conditional GWAS with Chr8 top SNP | | Conditional GWAS with Chr14 top SNP | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SNP | Chr | BP | Alleles | Case | Control | $R^2$ from top SNP | GWAS results OR | GWAS results P-value | OR | P-value | OR | P-value |
| BICF2S23237035 | 2 | 77,686,623 | T/C | 0.11 | 0.05 | | 1.41 | 6.21E-05 | | | | |
| BICF2P1011303 | 8 | 52,650,576 | T/C | 0.18 | 0.07 | 0.53 | 1.30 | 4.60E-05 | 1.03 | 0.663 | | |
| BICF2P29000 | 8 | 52,763,337 | C/A | 0.25 | 0.12 | 0.28 | 1.24 | 7.46E-05 | 1.07 | 0.170 | | |
| BICF2P378684 | 8 | 53,742,667 | C/T | 0.15 | 0.04 | 0.59 | 1.42 | 5.71E-06 | 1.04 | 0.560 | | |
| BICF2P1080535 | 8 | 53,778,185 | T/C | 0.16 | 0.05 | 0.75 | 1.36 | 1.50E-05 | 0.99 | 0.935 | | |
| BICF2P1048848 | 8 | 53,785,948 | A/G | 0.16 | 0.05 | 0.75 | 1.36 | 1.50E-05 | 0.99 | 0.935 | | |
| BICF2P184533 | 8 | 53,796,442 | G/A | 0.16 | 0.06 | 0.79 | 1.37 | 7.36E-06 | 1.00 | 0.978 | | |
| **BICF2P948919** | 8 | 53,818,371 | G/A | 0.21 | 0.07 | 1.00 | 1.39 | 2.66E-07 | 1.00 | 1.000 | | |
| TIGRP2P186605 | 14 | 11,778,977 | G/A | 0.66 | 0.43 | 0.93 | 1.18 | 5.13E-05 | | | 1.00 | 0.995 |
| BICF2G630521678 | 14 | 11,791,385 | A/G | 0.67 | 0.43 | 0.99 | 1.18 | 3.00E-05 | | | 1.00 | 0.966 |
| **BICF2G630521681** | 14 | 11,794,735 | C/T | 0.67 | 0.43 | 1.00 | 1.19 | 2.28E-05 | | | 1.00 | 1.000 |
| BICF2G630521696 | 14 | 11,807,161 | G/A | 0.67 | 0.43 | 0.99 | 1.18 | 3.67E-05 | | | 1.00 | 0.934 |
| BICF2S23029378 | 17 | 4,217,272 | G/A | 0.88 | 0.73 | | 1.19 | 9.47E-05 | | | | |
| BICF2G630222435 | 17 | 8,102,574 | T/G | 0.83 | 0.64 | | 1.18 | 8.66E-05 | | | | |
| BICF2P916139 | 17 | 8,135,932 | T/A | 0.82 | 0.63 | | 1.18 | 8.39E-05 | | | | |
| BICF2G630221951 | 17 | 8,819,612 | C/T | 0.85 | 0.65 | | 1.19 | 9.42E-05 | | | | |
| BICF2P780894 | 29 | 10,587,617 | G/A | 0.63 | 0.46 | | 1.18 | 7.45E-05 | | | | |

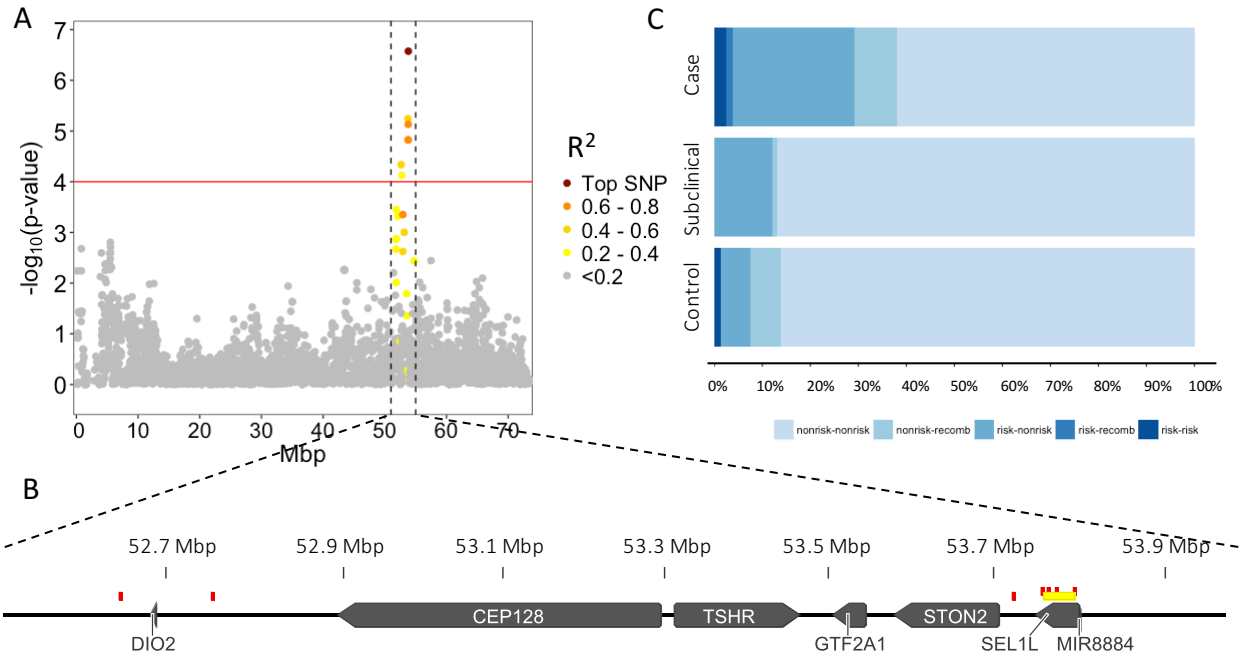



**Figure 3.2. Close-up of the chromosome 8 peak.** (A) $R^2$ from top SNP (BICF2P948919) is depicted to show LD structure. (B) Close-up view of the genes located in the region with $R^2$>0.2. All associated SNPs are depicted in red; the haplotype block containing the top 4 SNPs is highlighted in yellow. (C) Haplotype block containing the 4 associated SNPs (BICF2P1080535, BICF2P1048848, BICF2P184533, and BICF2P948919). The risk haplotype was TAGG and non-risk was CGAA. Dogs were considered recombined if neither combination was present.

*Shared association with mast cell tumor cases on chromosome 14*

The second top association peak is on chromosome 14 and contains four SNPs from 11,778,977–11,807,161 bp (Table 3.1). All SNPs were in strong LD ($R^2$>0.9) with the top SNP (OR = 1.18, p-value = $8.39 \times 10^{-5}$). Three of the four SNPs had previously been reported to be associated with mast cell tumors (MCTs) among American Golden Retrievers.[128] Thus, we assessed our data in combination with the American Golden Retriever data from the publicly available MCT dataset. Files were merged so that the new "case" population included TZL cases and MCT cases, whereas the reference population contained subclinicals and controls from the TZL dataset and controls from the MCT dataset. The chromosome 14 peak for the

combined dataset was wider and more strongly associated, with the top SNP reaching p = $1.5x10^{-9}$ (Figure 3.3). A GWAS including the TZL dataset and only MCT controls showed no increased association at the chromosome 14 peak (Suppl. Fig. 3.4B), confirming that this is a shared association for the two different cancers and not simply a result of increased power from the additional controls. We evaluated haplotype blocks in the combined dataset. The top SNP from the combined dataset was the same as the top SNP in the TZL only dataset (BICF2G630521681; Table 3.2). These SNPs are part of a ten SNP haplotype block that spans 11,695,969 – 11,807,161 bp (Figure 3.4). When we ran a conditional GWAS controlling for the top SNP, none of the SNPs in the larger associated region remained significant, suggesting they all represent one signal (Table 3.2). The haplotype block containing the top SNP spans three hyaluronidase enzymes, including Sperm Adhesion Molecule 1 (SPAM1; formerly called HYAL1), Hyaluronoglucosaminidase 4 (HYAL4), and a hyaluronidase 4-like gene (ENSCAFG00000024436). In our dataset, 85% of cases (67/79) had at least one risk haplotype (vs. 71% of subclinicals [77/108] and 65% of controls [52/80]); 18% of cases were homozygous (14/79) for the risk haplotype (vs. 7% of subclinicals [11/108] and 9% of controls [7/80]) (Figure 3.4).

*Targeted resequencing of chromosome 14 region identifies coding mutations in hyaluronidase genes*

Median coverage across the 8Mb region on chromosome 14 was 140x; 1,404 SNVs and 742 indels passed our filters. Five mutations causing amino acid changes were identified within coding regions of the three hyaluronidase genes (SPAM1, HYAL4, and ENSCAFG00000024436) (Figure 3.4); all mutations followed the GWAS haplotype. The mutation within the SPAM1 gene (cfa14:11,704,952, Lys482Arg) was predicted to be "possibly damaging" (PolyPhen-2 score 0.91).

The three mutations in the HYAL4 gene (cfa14:11,736,613, Gly454Ser; cfa14:11,736,674, Ser434Phe; cfa14:11,736,843, Leu378Ile) and one within ENSCAFG00000024436 (cfa14:11,760,826, Met463Thr) were predicted to be benign (PolyPhen-2 score <0.15). A LiftOver of these coordinates to CanFam2 determined the non-synonymous mutations in SPAM1 and HYAL4 were identical to those identified in the MCT study.[128] Additional non-coding variants were identified near these genes, including 5′ UTR variants (two SNVs, one indel in HYAL4), 3′ UTR variants (two SNVs, two indels in HYAL4 and three SNVs in SPAM1), up- and downstream gene variants, and intron variants (Suppl. Tables 3.3 and 3.4). One synonymous coding SNP was identified in ENSCAFG00000024436 (cfa14:11,768,664) (Suppl. Table 3.3).

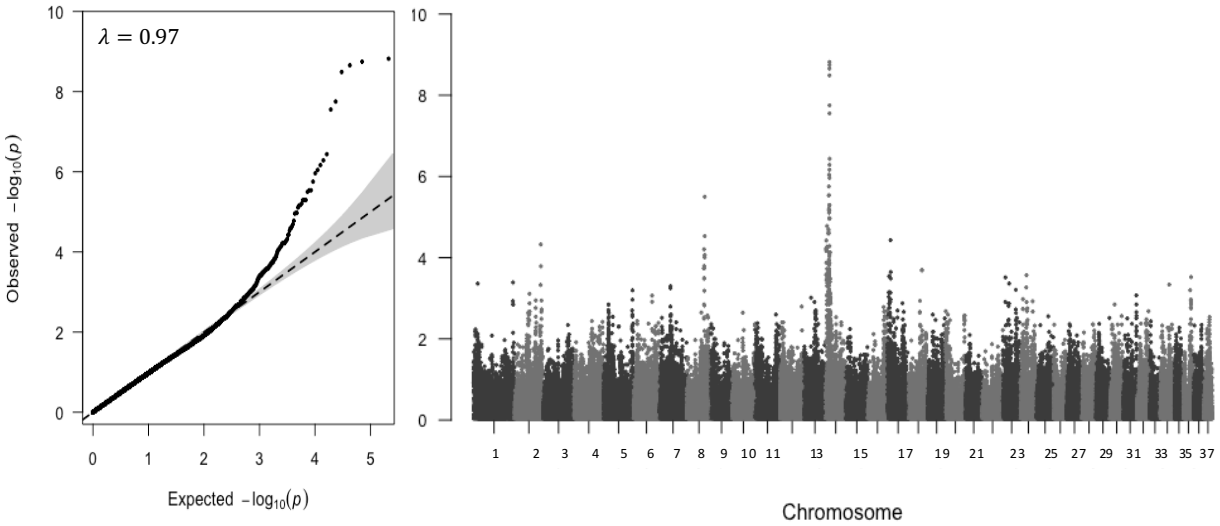



**Figure 3.3. GWAS results for combined TZL and MCT datasets.** QQ plot (left) and Manhattan plot (right).

**Table 3.2. SNPs in the chromosome 14 peak from the combined TZL + MCT GWAS.** SNPs contained in the haplotype block are shaded in grey; those shown in bold are also significant in the TZL only GWAS. The top SNP (smallest p-value) is shown in red.

| | | | | Associated Allele Freq | | $R^2$ from | TZL + MCT GWAS | | Conditional for Chr14 | |
|---|---|---|---|---|---|---|---|---|---|---|
| SNP | Chr | BP | Alleles | Case | Contr | top | OR | P-value | OR | P-value |
| BICF2G630517340 | 14 | 2,191,219 | T/G | 0.75 | 0.57 | 0.56 | 1.16 | 6.05E-05 | 0.98 | 0.6222 |
| BICF2G630517370 | 14 | 2,213,041 | A/G | 0.75 | 0.57 | 0.56 | 1.16 | 6.05E-05 | 0.98 | 0.6222 |
| BICF2G630517387 | 14 | 2,224,605 | T/C | 0.75 | 0.57 | 0.56 | 1.16 | 6.05E-05 | 0.98 | 0.6222 |
| BICF2G630517402 | 14 | 2,248,006 | T/C | 0.75 | 0.57 | 0.57 | 1.15 | 6.83E-05 | 0.98 | 0.6080 |
| BICF2G630517422 | 14 | 2,263,592 | A/G | 0.75 | 0.57 | 0.56 | 1.16 | 6.05E-05 | 0.98 | 0.6222 |
| BICF2G630517453 | 14 | 2,291,305 | G/A | 0.75 | 0.57 | 0.56 | 1.16 | 3.79E-05 | 0.99 | 0.7095 |
| BICF2G630519109 | 14 | 4,790,093 | G/A | 0.77 | 0.59 | 0.53 | 1.18 | 1.66E-05 | 1.00 | 0.9836 |
| BICF2G630519137 | 14 | 4,873,671 | G/T | 0.76 | 0.59 | 0.53 | 1.16 | 5.71E-05 | 0.99 | 0.7416 |
| BICF2P875875 | 14 | 5,198,586 | C/G | 0.89 | 0.75 | 0.31 | 1.19 | 5.32E-05 | 1.03 | 0.4517 |
| BICF2G630519935 | 14 | 9,007,123 | C/G | 0.77 | 0.60 | 0.42 | 1.16 | 2.04E-05 | 1.01 | 0.7179 |
| BICF2G630519944 | 14 | 9,028,362 | C/T | 0.77 | 0.60 | 0.42 | 1.16 | 2.55E-05 | 1.01 | 0.7536 |
| BICF2G630519947 | 14 | 9,054,201 | T/C | 0.77 | 0.60 | 0.42 | 1.16 | 2.55E-05 | 1.01 | 0.7536 |
| BICF2G630519948 | 14 | 9,061,567 | C/A | 0.77 | 0.60 | 0.42 | 1.16 | 2.28E-05 | 1.01 | 0.7240 |
| BICF2G630520440 | 14 | 9,897,520 | G/A | 0.70 | 0.49 | 0.66 | 1.18 | 2.91E-06 | 1.00 | 0.8889 |
| BICF2G630520475 | 14 | 9,910,392 | T/G | 0.70 | 0.49 | 0.67 | 1.18 | 6.23E-06 | 0.99 | 0.7312 |
| BICF2G630520518 | 14 | 9,930,996 | G/A | 0.70 | 0.49 | 0.66 | 1.18 | 2.91E-06 | 1.00 | 0.8889 |
| BICF2G630520533 | 14 | 9,944,218 | C/T | 0.70 | 0.49 | 0.67 | 1.19 | 1.76E-06 | 1.00 | 0.9524 |
| BICF2G630520585 | 14 | 9,976,888 | T/C | 0.72 | 0.53 | 0.61 | 1.18 | 6.95E-06 | 0.99 | 0.8432 |
| BICF2S23335956 | 14 | 10,450,765 | C/A | 0.62 | 0.39 | 0.72 | 1.19 | 9.05E-07 | 1.00 | 0.9119 |
| BICF2S23442141 | 14 | 10,462,378 | G/A | 0.63 | 0.43 | 0.64 | 1.18 | 7.81E-06 | 0.99 | 0.8229 |
| BICF2S2358168 | 14 | 10,997,271 | A/G | 0.90 | 0.77 | 0.27 | 1.20 | 7.88E-05 | 1.04 | 0.3753 |
| BICF2P546873 | 14 | 11,370,040 | G/A | 0.57 | 0.40 | 0.57 | 1.15 | 9.29E-05 | 0.98 | 0.5126 |
| BICF2P753712 | 14 | 11,396,405 | T/G | 0.55 | 0.35 | 0.67 | 1.18 | 5.08E-06 | 0.99 | 0.7453 |
| BICF2P1501412 | 14 | 11,406,182 | T/G | 0.56 | 0.36 | 0.67 | 1.17 | 1.10E-05 | 0.98 | 0.5800 |
| BICF2P1022612 | 14 | 11,425,262 | G/A | 0.55 | 0.35 | 0.67 | 1.18 | 5.08E-06 | 0.99 | 0.7453 |
| BICF2P448523 | 14 | 11,444,087 | T/C | 0.55 | 0.35 | 0.67 | 1.18 | 5.08E-06 | 0.99 | 0.7453 |
| BICF2P1199126 | 14 | 11,451,260 | T/G | 0.55 | 0.35 | 0.67 | 1.18 | 6.83E-06 | 0.99 | 0.6929 |
| BICF2G630521558 | 14 | 11,695,969 | C/T | 0.78 | 0.59 | 0.62 | 1.19 | 1.09E-06 | 1.00 | 0.9482 |
| BICF2G630521572 | 14 | 11,721,433 | T/C | 0.70 | 0.45 | 0.94 | 1.22 | 2.80E-08 | 0.99 | 0.7588 |
| BICF2G630521606 | 14 | 11,733,161 | T/C | 0.78 | 0.58 | 0.61 | 1.19 | 6.84E-07 | 1.01 | 0.8209 |
| BICF2G630521619 | 14 | 11,736,615 | C/T | 0.79 | 0.58 | 0.60 | 1.19 | 5.17E-07 | 1.01 | 0.7721 |
| BICF2P867665 | 14 | 11,765,081 | G/T | 0.79 | 0.57 | 0.59 | 1.22 | 1.77E-08 | 1.03 | 0.3593 |
| **TIGRP2P186605** | 14 | 11,778,977 | G/A | 0.70 | 0.43 | 0.95 | 1.23 | 3.25E-09 | 1.00 | 0.9681 |
| **BICF2G630521678** | 14 | 11,791,385 | A/G | 0.70 | 0.44 | 0.99 | 1.23 | 2.21E-09 | 1.00 | 0.9637 |
| **BICF2G630521681** | 14 | 11,794,735 | C/T | 0.70 | 0.43 | 1.00 | 1.23 | 1.51E-09 | 1.00 | 1.0000 |
| **BICF2G630521696** | 14 | 11,807,161 | G/A | 0.71 | 0.44 | 0.98 | 1.23 | 1.79E-09 | 1.00 | 0.9863 |
| BICF2P665918 | 14 | 11,852,463 | T/C | 0.69 | 0.53 | 0.38 | 1.18 | 1.07E-05 | 1.02 | 0.5686 |
| TIGRP2P186621 | 14 | 11,867,284 | A/G | 0.69 | 0.53 | 0.38 | 1.18 | 1.07E-05 | 1.02 | 0.5686 |
| BICF2G630521759 | 14 | 11,958,198 | T/C | 0.86 | 0.72 | 0.29 | 1.18 | 7.49E-05 | 1.03 | 0.4469 |
| BICF2P626537 | 14 | 12,061,360 | A/G | 0.38 | 0.23 | 0.30 | 1.17 | 5.20E-05 | 1.03 | 0.4442 |
| BICF2G630522103 | 14 | 12,249,856 | C/T | 0.39 | 0.24 | 0.30 | 1.16 | 8.23E-05 | 1.02 | 0.5203 |
| BICF2G630522165 | 14 | 12,431,638 | C/A | 0.81 | 0.63 | 0.45 | 1.22 | 3.65E-07 | 1.03 | 0.3555 |
| BICF2G630522789 | 14 | 13,423,322 | C/T | 0.59 | 0.42 | 0.48 | 1.16 | 7.39E-05 | 1.00 | 0.8819 |

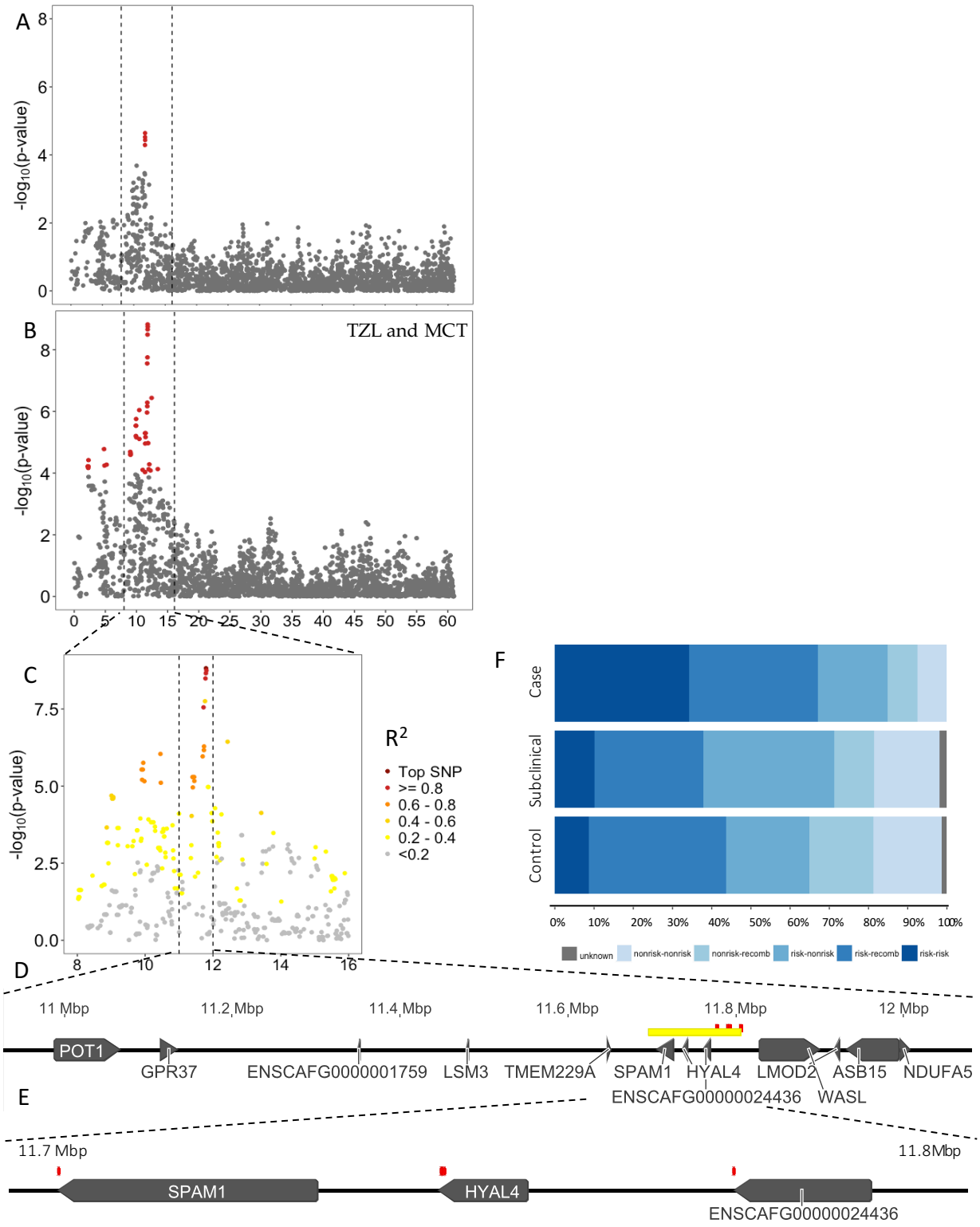



**Figure 3.4. Close-up of the chromosome 14 peak depicting change in signal with MCT dataset added.** (A) TZL dataset only. (B) Combined TZL and MCT dataset; (C) Close-up of the region from 8-12Mbp containing SNPs with $R^2$ >0.2. (D) Close-up view of genes located in the region from 11-12Mb. The 4 SNPs significantly associated with TZL are depicted in red and the ten-

SNP haplotype block they represent is shaded in yellow. (E) Close-up of the region from 11.7-11.8 Mbp where coding mutations (shown in red) were found on resequencing. (F) Haplotype block containing 10 associated SNPs on cfa14 (BICF2G630521558, BICF2G630521561, BICF2G630521572, BICF2G630521606, BICF2G630521619, BICF2P867665, TIGRP2P186605, BICF2G630521678, BICF2G630521681, BICF2G630521696). The risk haplotype was CTTTCGGACG and non-risk was TTCCTTAGTA. Dogs were considered recombined if neither combination was present and were considered unknown if the genotype for one or more SNPs was missing.

*Potential cumulative risk for chromosomes 8 and 14*

Distribution of haplotype scores are shown in Figure 3.5. A higher proportion of cases (53% vs. 19% of subclinicals and 14% of controls) had a haplotype score ≥2, indicating a potential cumulative risk based on number of risk haplotypes. With the exception of one subclinical dog, only cases had ≥3 risk haplotypes. Larger sample sizes are necessary to evaluate statistical interaction of the chromosome 8 and 14 haplotypes.

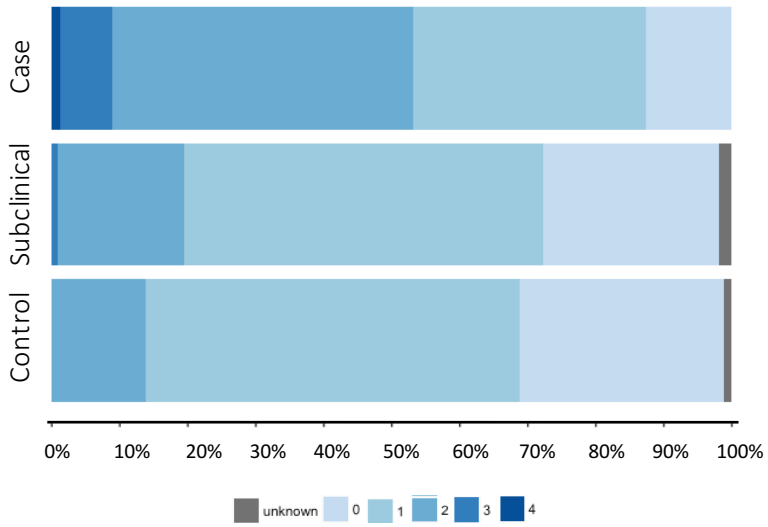



**Figure 3.5. Distribution of haplotype scores.** Dogs were scored from zero to four based on the number of risk haplotypes for chromosomes 8 and 14. Recombined haplotypes were considered non-risk. Dogs were considered unknown if the genotype for one or more SNPs was missing.

*Additional significantly associated GWAS SNPs*

Associated SNPs were also seen on chromosomes 2, 17, and 29, but our study did not have the power to accurately determine the regions of association. We conducted a restricted maximum likelihood analysis, assuming TZL has a 2% prevalence in the Golden Retriever breed, and found that the combined set of 17 significant SNPs in our dataset explained approximately 15% (standard error 7%) of the phenotypic variance, whereas all GWAS SNPs explained approximately 49% (standard error 13%).

**Discussion**

As Golden Retrievers are overrepresented among dogs with TZL, we were able to conduct a genome-wide association study to identify genetic risk factors for TZL in this breed. We identified associated regions on both chromosome 8 and 14. Subsequent resequencing of a subset of dogs identified non-synonymous mutations in three hyaluronidase genes on chromosome 14. Coding mutations were not found in the chromosome 8 region, but identified variants may be located in regulatory elements for numerous genes, including DIO2, CEP128, GTF2A1, STON2, and SEL1L.

*Mutations in hyaluronidase genes associated with risk for TZL and MCT*

GWAS analysis and subsequent resequencing identified mutations in SPAM1 and HYAL4 identical to those seen in Arendt et al.'s MCT study,[128] highlighting a potential shared mechanism for TZL and MCT pathogenesis. One potential mechanism is via hyaluronan turnover, which is caused by the interaction of hyaluronan and CD44, a cell surface

glycoprotein expressed on both T cells and mast cells.[139] This turnover leads to increased low molecular weight hyaluronan, the byproducts of which are pro-inflammatory and pro-oncogenic, with implications in cell proliferation, migration, and angiogenesis.[140, 141] In contrast, high molecular weight hyaluronan and decreased hyaluronidase activity have been associated with the increased longevity and cancer resistance seen in naked mole rats.[140] It would be informative to measure hyaluronan in cases and controls to determine whether the ratio of low to high molecular weight hyaluronan is altered in TZL.

Most mammals have six hyaluronidase-like genes, clustered on two chromosomes. In dogs, HYAL1, HYAL2 and HYAL3 are clustered on cfa20, whereas SPAM1, HYAL4, and ENSCAFG00000024436 are clustered on cfa14. ENSCAFG00000024436 is homologous to HYALP1, which is an expressed pseudogene in people.[142] Hyalp1 is believed to be functional in other mammals,[142] although its functional status is unknown in dogs. Spam1 is considered a testis hyaluronidase and is important during egg fertilization by sperm.[143] However, it has been detected in the epididymis, seminal vesicles, prostate, female genital tract, breast, placenta, fetal tissue, and certain malignancies,[144-146] suggesting it is multifunctional and not sperm-specific. Despite the potential shared pathogenesis of TZL and MCT, we did not see an association between MCT and TZL in our dataset. Of dogs where medical history was known, 3/76 TZL cases (4%), 8/142 subclinicals (6%), and 4/103 controls (4%) had a history of or concurrent MCT. This suggests these diseases develop independently despite their shared mechanisms. More research is necessary to understand the role of these hyaluronidases in dogs and evaluate how the observed variants alter expression of hyaluronidases and downstream signaling.

*Potential role of vesicle trafficking in lymphoma predisposition*

Since we do not know the functional impact of variants in the cfa8 region, it is difficult to predict what role they play in TZL pathogenesis. Several genes are plausible candidates, and it is possible that regulatory variants affect the expression of multiple genes. As such, we can hypothesize many potential mechanisms. SEL1L, for example, encodes a protein that is part of a complex involved in endoplasmic reticulum-associated degradation of misfolded proteins.[13] If this complex cannot form correctly, accumulation of misfolded proteins can lead to endoplasmic reticulum stress, which has been proposed as a mechanism underlying diabetes mellitus,[147] rheumatoid arthritis,[148] and neurodegenerative disease.[149, 150] Additionally, SEL1L is a regulator of the Notch signaling pathway, which plays a role in cell differentiation, proliferation, and apoptosis.[151] Some researchers believe this gene acts as a tumor suppressor, as its downregulation has been associated with poor prognosis in pancreatic cancer pathogenesis,[152] breast carcinoma[153] and glioma.[154]

Another feasible mechanism is via the role of vesicle trafficking in antigen processing and presentation. When antigen-presenting cells encounter antigen, they internalize the antigen (if it is extracellular), process it, and present it to T cells on major histocompatibility complex.[155] This process involves several pathways, including receptor-mediated endocytosis, for which STON2 encodes a protein.[14, 156] Additionally, dendritic cells use the endoplasmic reticulum-associated degradation pathway during antigen cross-presentation.[157-159] Less is known about CEP128, and further research is needed to understand its role in normal animals. Alterations in the ability to present antigen could have important implications in T cell activation and immune response. No prior research has found associations of germline changes in STON2 and cancer,

but a recent study found somatic CEP128 mutations in a case of refractory diffuse large B cell lymphoma.[160]

*Thyroid hormone metabolism may influence TZL risk*

In a parallel study, we determined dogs with hypothyroidism were significantly less likely to get TZL as dogs without hypothyroidism. As thyroid hormone plays an important role in cell growth and metabolism, we hypothesize that lack of this hormone may decrease T cell proliferation and therefore help prevent the development of TZL. While SnpEff did not predict our SNVs to be modifiers of DIO2 or TSHR, it is possible that the regulatory elements of these genes are far up- or downstream, as seen in people. Canine genome annotations for this region may not yet be able to predict these relationships. While canine hypothyroidism is generally thought to be caused by lymphocytic thyroiditis or idiopathic atrophy,[161] it is plausible that changes in expression of DIO2 or TSHR could influence its development. Thyroid hormone regulation depends on an axis of multiple hormones and organs. Thyroid stimulating hormone, released from the pituitary, binds TSHR on the thyroid gland, causing release of thyroxine and, to a lesser extent, triiodothyronine.[162] DIO2 is one of two hormones responsible for converting thyroxine to triiodothyronine, the more active form, in the peripheral organs.[163] It is feasible that changes in the expression of either of these genes could alter thyroid hormone production and function. Interestingly, levels of Deiodinase 2, the product of DIO2, are tightly regulated, and synthesis can be inhibited by endoplasmic reticulum stress via endoplasmic reticulum-associated degradation.[164] Thus, alterations in SEL1L and endoplasmic reticulum stress could also impact thyroid hormone regulation.

*Golden Retriever predisposition to TZL*

We believe subclinicals represent a precursor state to TZL, so we hypothesized they would share the same genetic variants as cases. However, we were unable to differentiate subclinicals from controls in our GWAS analysis. The high prevalence of TZL among Golden Retrievers and corresponding high prevalence of subclinicals suggest the genetic basis for developing $CD45^-$ T cells may be fixed among this breed. If this is the case, we would be unable to identify the genetic risk factor for developing $CD45^-$ T cells in our study. Future studies may evaluate fixed regions of the Golden Retriever genome to identify candidate genes. Additionally, a GWAS of TZL among a less predisposed breed may distinguish additional associated regions not identified in our study.

*Conclusions*

Little is known about the functional implications of the mutations identified on chromosome 8. Since variants in this region are in moderate to high LD, it is difficult to prioritize which variants are important in disease pathogenesis vs. which are bystanders inherited with the causative mutation. Additional studies are necessary to elucidate these associations and better understand the effect of these variants. The likely importance of hyaluronidases and shared association with MCT is noteworthy and warrants further investigation. Further research will increase our understanding of how these coding mutations alter hyaluronidase function. Ultimately, future studies will help elucidate TZL pathogenesis and identify causative variants that may be biomarkers or disease risk or potential therapeutic targets.

# CHAPTER 4: AIM 3 – EXAMINE THE ASSOCATION OF CHRONIC INFLAMMATORY CONDITIONS AND T ZONE LYMPHOMA


**Summary**

*Background*: T zone lymphoma (TZL), a subtype of peripheral T cell lymphoma, accounts for about 12% of all canine lymphomas. Previously, we found that approximately 30% of Golden Retrievers without lymphocytosis or lymphadenopathy have cells phenotypically similar to those seen in TZL present in their peripheral blood. These animals are referred to as "subclinical" and present a unique opportunity to evaluate the etiology and pathogenesis of TZL through an epidemiologic analysis of cases, subclinicals, and controls. Since TZL has an activated phenotype, we hypothesize this disease is antigen-driven and a persistent antigen stimulus is necessary for progression of TZL.

*Objective*: To evaluate the association of chronic inflammatory conditions and TZL using a case-control study of Golden Retriever dogs.

*Methods*: Cases from throughout the United States were recruited through Colorado State University's Clinical Immunology laboratory, which diagnoses lymphoproliferative disorders using immunophenotyping. Our reference population was recruited through 1) a database of owners interested in participating in research studies and 2) the submitting clinics of cases. Dogs were classified as controls or subclinicals via flow cytometry of peripheral blood samples. Data on health history, signalment, environmental exposures and lifestyle were obtained from owner-completed questionnaires. ORs and 95% CIs were estimated using multivariable logistic regression.

*Results*: The study population included 140 cases, 221 subclinicals, and 147 control dogs. We identified two shared risk factors for cases and subclinicals, including history or concurrent bladder infection (OR 3.5, 95% CI 1.0–12.7 for cases; OR 5.1, 95% CI 1.9–13.7 for subclinicals) and eye disease (OR 2.3, 95% CI 1.0–5.2 for cases; OR 1.9, 95% CI 1.0–3.8 for subclinicals). Mange and gastrointestinal disease were also significantly associated with TZL cases (OR 5.5, 95% CI 1.4–21.1 and OR 2.4, 95% CI 1.0–5.8, respectively). Timing of spay/neuter was significantly associated with both cases and subclinicals, with late spay (>1 vs. <1 year of age) reflecting decreased odds of being a case (OR 0.3, 95% CI 0.2–0.7), whereas late neuter was associated with increased odds of being subclinical (OR 2.2, 95% CI 1.1–4.4). In addition, use of omega-3 supplements and hypothyroidism were associated with decreased odds of being a case (OR 0.3, 95% CI 0.1–0.6 and OR 0.3, 95% CI 0.1–0.7, respectively).

*Conclusions*: We identified several factors associated with inflammation (mange, bladder infections, eye disease, gastrointestinal disease) that are positively associated with TZL, and one factor associated with dampening inflammation, the use of omega-3 supplements, that is protective. These findings are consistent with the activated phenotype of the tumor cells, and support the hypothesis that inflammation is a risk factor for TZL. We also found hypothyroidism was associated with a decreased TZL risk; further research is needed to understand the mechanism underlying this association.

**Introduction**

Peripheral T cell lymphomas (PTCLs) are a heterogeneous group of malignancies derived from mature T cells and NK cells.[20] The etiology of most PTCL subtypes is poorly understood, largely due to their low incidence in people, which makes it difficult to study subtypes independently. T zone lymphoma (TZL), an indolent subtype of PTCL, accounts for approximately 12% of all canine lymphomas.[40] TZL can be readily diagnosed through identification of a homogeneous expansion of T cells lacking expression of the cell surface marker CD45 using flow cytometry.[6] This disease appears to have a striking predilection for Golden Retrievers, with over 40% of cases occurring in this breed.[6, 28] This observation suggests a genetic component of this multifactorial disease. We have observed that over 30% of Golden Retrievers without a lymphocytosis or lymphadenopathy have $CD45^-$ T cells in their blood present at "subclinical" levels (hereafter referred to as "subclinical" dogs) (Hughes et al., under review). The lack of clinical signs suggests that these are pre-neoplastic T cells with an unknown potential for progressing to overt neoplasia. Because lack of CD45 expression may be associated with decreased apoptosis, it is possible that antigen stimulation could lead to $CD45^-$ T cells undergoing unregulated division.[165]

Multiple studies have provided evidence for the plausibility of chronic antigen stimulus playing a role in the pathogenesis of TZL. First, TZL cells have an activated phenotype, expressing high levels of CD21 and CD25,[6] suggesting the tumor arose from a T cell that had been activated by antigen. Second, human epidemiologic evidence suggests autoimmune disease and allergy are associated with PTCL.[63, 103] Third, a recent murine study[5] of PTCL showed cells have a chronically activated phenotype and demonstrated reduced tumor growth

by blocking antigen binding to the T cells. Thus, we suspect chronic inflammatory diseases will persistently stimulate CD45$^-$ T cells to proliferate, predisposing them to progress to clinically apparent TZL. As such, the objective of this study was to evaluate the association of chronic inflammatory conditions and TZL using a case-control study of Golden Retriever dogs. By utilizing a unique population of dogs including cases, subclinicals, and controls, we aim to provide insights into the etiology and pathogenesis of TZL.

## Methods

### *Study participants*

We conducted a case-control study, the Golden Years Study, through the Colorado State University Clinical Immunology (CSU-CI) laboratory from October 2013 through March 2016. Golden Retrievers with TZL (cases) were identified through submissions from clinics throughout the United States (Figure 4.1). Two series of lymphoma-free Golden Retrievers, aged 9 years and older, were recruited as our comparison group: database-recruited and clinic-recruited. *Database-recruited* dogs were enrolled through email solicitation to a database of Golden Retriever owners who were interested in participating in research studies. The database was developed as part of the Canine Lifetime Health Project,[32] which allowed owners to register to receive information about studies they may be eligible for by providing their dog's breed and age. *Clinic-recruited* dogs were enrolled from the submitting clinic of cases; veterinarians were asked to enroll their next 1-2 Golden Retriever patients that met our enrollment criteria. Dogs in the comparison group were subsequently categorized as subclinicals or controls using flow cytometry, as described below.

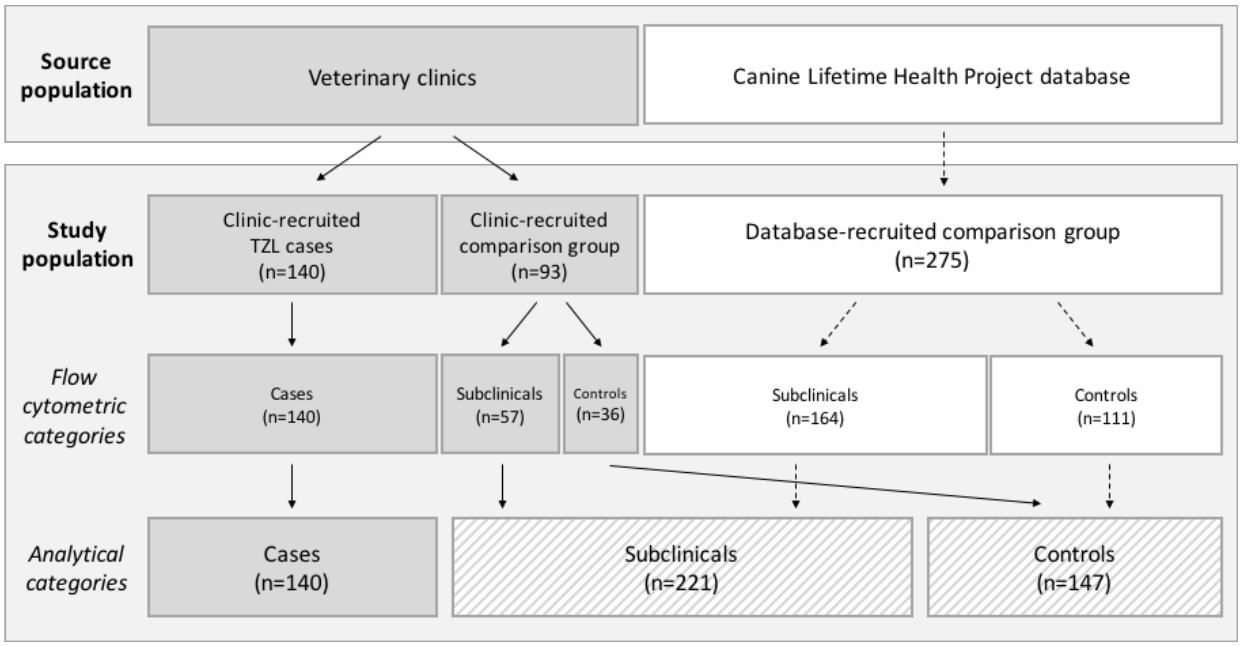



**Figure 4.1. Diagram of recruitment and study population.** TZL cases were recruited from veterinary clinics throughout the U.S. Lymphoma-free Golden Retrievers aged 9 years and older were recruited from both veterinary clinics and the Canine Lifetime Health Project database as the comparison group. All dogs recruited for the comparison group were tested via flow cytometry for designation as either subclinical or control dogs. Recruitment strategies were subsequently combined based on flow cytometric categorization to create 3 groups: cases, subclinicals, and controls.

*Case and control definition*

Flow cytometric analysis of peripheral blood samples was used to diagnose dogs as cases, subclinicals, or controls. Flow cytometry was carried out as previously described[6] and samples were analyzed with the antibody combinations listed in Suppl. Table 3.1 using a 3-laser Coulter Gallios (Beckman Coulter Inc., Brea, CA). All data analyses were conducted using Kaluza Analysis Software (Beckman Coulter Inc., Brea, CA). *Case* definition included a homogeneous expansion of $CD5^{+}CD45^{-}$ T cells as well as the presence of lymphocytosis (defined as >5,000 lymphocytes/μL) and/or lymphadenopathy (Suppl. Figure 3.1A). *Controls* had no history or clinical signs of a LPD (including no lymphadenopathy or lymphocytosis) and had no

population of $CD5^{+}CD45^{-}$ T cells identified by flow cytometry (Suppl. Figure 3.1C). *Subclinicals* also had no history or clinical signs of a LPD, but had a small population of $CD5^{+}CD45^{-}$ T cells on flow cytometry (>1% of total T cells) without the presence of a lymphocytosis or lymphadenopathy (Suppl. Figure 3.1B). Our desired sample size was calculated to detect an OR of 1.5-2.5, assuming exposure percentage of 30-50% among controls, a 2:1 ratio of controls to cases, an alpha of 0.05, and 80% power. Using these parameters, we calculated an estimated 200 cases and 400 controls to detect an OR of 1.7, and as few as 80 cases and 160 controls to detect an OR >2.0. Since the population of subclinical dogs was discovered while conducting this study, they were not considered in our power calculations.

*Data collection and exposure assessment*

We developed a broad-based questionnaire, based on the Golden Retriever Lifetime Study questionnaire,[32] to capture potential TZL risk factors and gain a better understanding of our population demographics. Modifications to the existing questionnaire included altering phrasing to capture the dog's lifetime (vs. annual updates) and adjusting language about medications and health history for a lay audience (vs. veterinarian). The resulting questionnaire included 44 multi-part questions and was implemented online through SurveyMonkey® (Suppl. Figure 4.2). Owners were instructed to complete the questionnaire within three months of submission of the blood sample. Once initiated, it took an average of 14 minutes to complete. Information ascertained included the dog's health history (e.g. cancer, infectious diseases), environmental exposures (e.g. time in rural areas, exposure to smoke), medications (e.g. antihistamines, steroids, anti-inflammatories), signalment (age, sex, timing of spay/neuter, zip

code of residence), preventive care (flea/heartworm preventives, dental care, vaccinations), and diet (Suppl. Figure 4.1). All disease history information was obtained based on age at diagnosis (categorized as <1 year of age, 1-3 years, 4-6 years, 7-10 years, or >10 years), but was collapsed into ever/never diagnosed due to sparse data within categories. Health history variables were left as broad categories unless we had a specific hypothesis to evaluate a sub-category.

Health history

Specific risk factors of interest included infectious diseases, allergic disorders, autoimmune diseases, and inflammatory disorders that could cause chronic immune stimulation. Among infectious diseases, we were particularly interested in history of parasitic or nematode infection due to preliminary evidence from our lab that suggests TZL cells may be of Th2 origin. This information was gathered as a history of "worms" in the questionnaire. All other infectious diseases were collapsed into a single category. We evaluated bladder infections separately from other infectious diseases because information about them was ascertained within the category of urinary and reproductive conditions. Likewise, ear infections were evaluated separately because they were ascertained among eye, ear, nose, and throat conditions. Allergic and autoimmune disorders of interest included skin disease, vaccine reactions and hypothyroidism. Due to the likely misclassification of types of skin diseases using owner report, all skin diseases except mange were collapsed into one variable. A previous report[28] indicated approximately 10% of TZL cases present with demodectic mange, leading to an *a priori* decision to evaluate mange separately from other skin diseases. Inflammatory disorders of interest included dental disease, eye disease, cardiovascular disease, gastrointestinal disease, and orthopedic conditions. Uveitis was evaluated separately from other eye diseases because of the

strong genetic component of this disease.[166-168] In addition, orthopedic disease was split into cruciate ligament rupture and "degenerative joint disease," which included elbow and hip dysplasia, intervertebral disc disease, and osteoarthritis, due to differences in underlying diseases processes.

Environmental exposures

Environmental exposures were also of interest because of the hypothesized importance of immune stimulation in the etiology of TZL. Exposure to rural environments, parks, lawn chemicals, cigarette smoke, and chemicals were ascertained as daily, weekly, monthly, occasionally, or never. Frequency of swimming in irrigation water/ponds/canals, lakes/streams, or the ocean was ascertained using the same categories. For simplicity, and to avoid sparse cells, these categories were collapsed as "frequent" vs. "infrequent", with frequent comprising daily, weekly, or monthly exposure and infrequent comprising occasional or null exposure.

Medications

Certain medications were of interest due to their ability to dampen the immune response, and therefore possibly reduce the risk of TZL. Our main categories of interest included antihistamines, non-steroidal anti-inflammatories, steroids, antibiotics, immunosuppressants, and non-prescription supplements. Supplementation with omega-3s was a hypothesized protective factor because of its anti-inflammatory properties,[169-171] so use of non-prescription supplements was split into omega-3 and "other".

Signalment

Sex and age at spay/neuter were of interest due to multiple recent papers indicating an association of early spay/neuter and lymphoma.[61, 172-174] Age at spay/neuter was categorized as

female spayed prior to 1 year of age (used as the reference), female spayed after 1 year of age (or intact), male neutered prior to 1 year of age, and male neutered after 1 year of age (or intact).

Due to our recruitment strategy, we could not evaluate associations of TZL with age or diagnoses of non-lymphoma cancers. Age was restricted to 9 years or older among the control and subclinical populations, but younger cases were enrolled in the study, which resulted in a difference in the age distribution for case and non-cases populations. In addition, there was an overrepresentation of non-lymphoma cancers among control and subclinical populations using both recruitment strategies. Since cases often came from cancer specialty clinics, the corresponding clinic-recruited non-cases had non-lymphoma cancers, thereby skewing the proportion of non-case dogs with cancers. In addition, many owners of database-recruited dogs originally thought only dogs with cancer were allowed to participate in the study, also leading to an overrepresentation of non-lymphoma cancers among this population.

Preventive care

We additionally asked numerous questions about preventive care in attempt to understand the demographics of our population. These questions included use of flea and heartworm preventives as well as vaccination frequency.

*Statistical Analysis*

Inclusion criteria were re-checked based on survey results, and non-case dogs whose owners indicated they had been diagnosed with lymphoma or were <9 years old were excluded. In addition, dogs were excluded if only the demographic portion of the survey was completed.

A "missing" category was created for variables missing data for >5% of observations (age at spay/neuter, Bordetella vaccination frequency, all medication use variables, exposure to parks and rural environment, swimming in the ocean, irrigation water, or lakes/streams); all other variables were binary. Data were summarized as frequency and percent or median and range as appropriate.

Modeling strategy

As no prior research has been conducted investigating risk factors for TZL, we used a data-driven approach to evaluate variables that fit our hypothesis. Univariable associations for case vs. control and subclinical vs. control were calculated using logistic regression. We chose to analyze subclinicals as a distinct group to evaluate whether their exposure levels were more similar to cases or controls, or were intermediate between those groups. Variables were selected for consideration for the multivariable model if they fell into any of the following categories: 1) univariably significant at $p<0.25$ (using type 3 analysis of effects), 2) variable of concern based on association with recruitment strategy (case vs. control analysis only), or 3) "reasonable" variance (i.e. not near-zero variance; using nearZeroVar function in R's caret package[175]). All variables that met the above criteria were put into a model and assessed for collinearity using variance decomposition proportions and condition indices[176] prior to performing best subsets regression (Figure 4.2). Indicator variables were created for best subsets regression. A consensus model was then built using the variables deemed important by best subsets as well as their required constituents (i.e. if only one indicator variable from a multi-level variable was chosen by best subsets, all levels were included in the consensus model). Our goal was to choose a consensus model with no more than one variable for every 10 participants in the smallest group

(i.e. no more than 14 variables [14 degrees of freedom] in a model assessing 140 cases). Variables were selected for the consensus model based on "tiers" until we reached the maximum degrees of freedom. For instance, if our maximum degrees of freedom was 14 and the best 8-variable model contained 10 degrees of freedom, we considered those variables "first tier", the variables from the best 10-variable model "second tier", and the variables from the best 12-variable model "third tier", and so on until 14 degrees of freedom were reached.

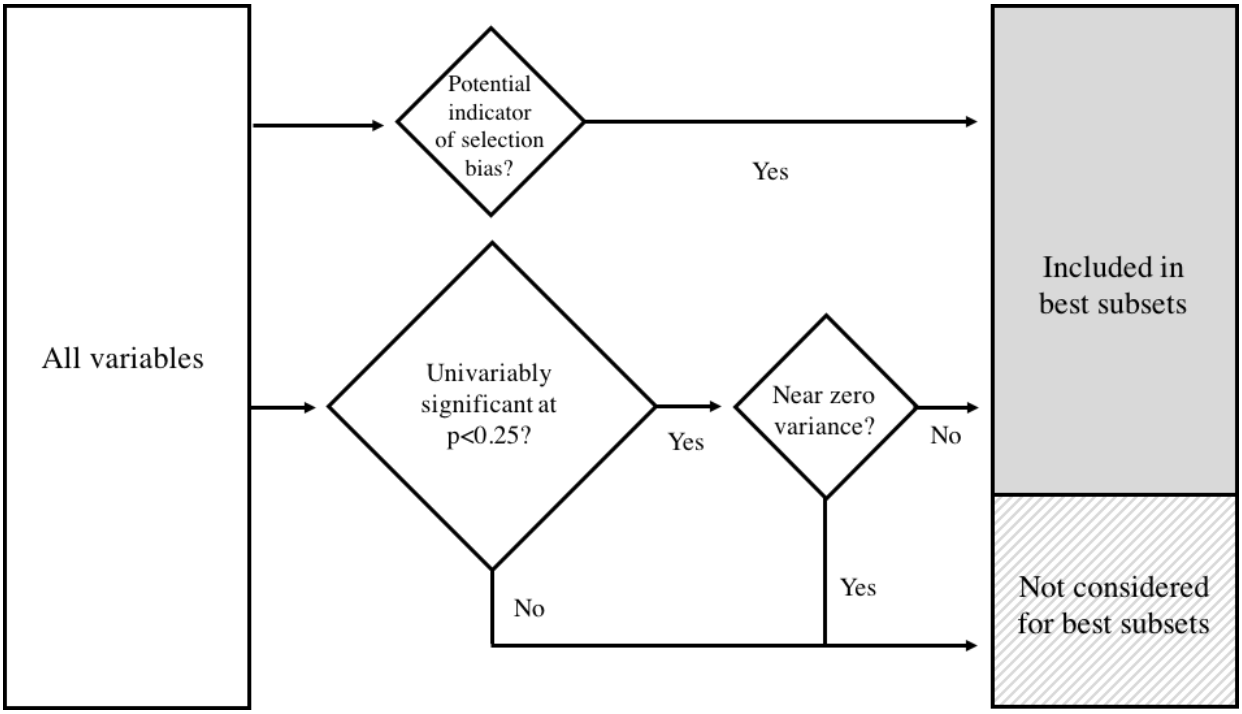



**Figure 4.2. Flow chart of variable selection for multivariable modeling.** All variables selected to be included in best subsets were evaluated for collinearity before proceeding with best subsets regression.

The consensus model was used as the "full" model and variables were eliminated in a stepwise manner.[176] At each step, 1) the variable with the highest, non-significant p-value (>0.05) was removed, 2) confounding was assessed based on a >10% change in OR estimates from the full model, and 3) a likelihood ratio test was used to compare the resulting model to the prior model. This process was repeated until all variables were significant at p<0.05, removal of a

variable changed at least one OR by >10%, or the likelihood ratio test suggested the prior model performed better than the reduced model.

Evaluation of recruitment bias

To determine whether there were systematic differences between database- and clinic-recruited dogs in the comparison group, we conducted univariable analyses to assess the association of recruitment strategy and each variable of interest among 1) subclinicals and 2) controls. Fisher's exact tests were calculated for all variables; variables significantly associated with recruitment strategy ($p<0.05$) among both controls and subclinicals were noted as potential indicators of recruitment bias for multivariable modeling. Recruitment strategy could not be used as an indicator variable in case vs. control analyses because cases were only recruited through clinics.

The concern for selection bias and rationale for including an indicator are shown in the directed acyclic graph (DAG) (Figure 4.3). By virtue of the case-control design, TZL status is associated with recruitment strategy. Thus, any exposure variable (labeled "Exposure 1" in Figure 4.3) associated with recruitment strategy would cause recruitment to become a collider. As such, there would be a collider bias in our analysis, imposing an association between TZL and exposure 1. This could bias the association of any other exposure variables (i.e. "Exposure 2" in Figure 4.3) that are associated with exposure 1. To block this biasing path, exposure 1 would need to be included in the model.

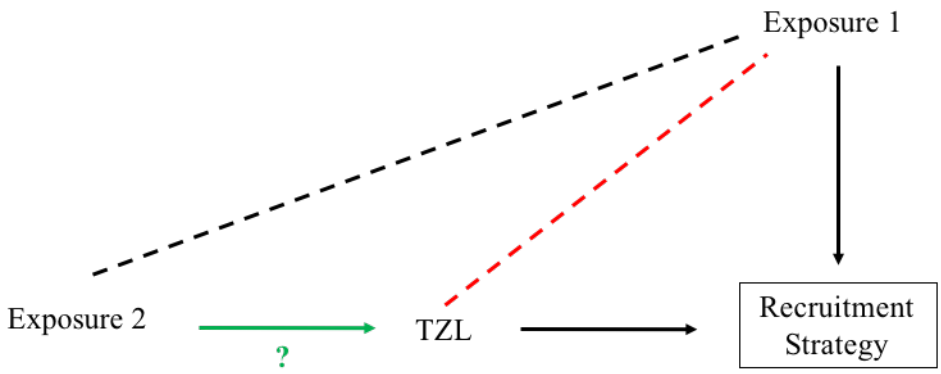



**Figure 4.3. Directed acyclic graph showing potential recruitment bias**. We are interested in measuring the association of exposure 2 and TZL (green arrow). If exposure 1 is associated with recruitment strategy in our study, when we condition on recruitment strategy (by analyzing dogs enrolled in our study), we induce an association of TZL and exposure 1. If exposure 1 and exposure 2 are also associated, we will have a biasing path from exposure 2 → exposure 1 → TZL.

Sensitivity analyses

For our final models, we conducted sensitivity analyses to evaluate the impact of age and non-lymphoma cancer on our results. First, we removed cases who were <9 years of age and re-ran the case vs. control final model. Second, we removed all dogs (cases, subclinicals, and controls) with diagnoses of non-lymphoma cancer and re-ran both case vs. control and subclinical vs. control models.

This study was conducted with approval from the Colorado State University (CSU) Institutional Animal Care and Use Committee. All data cleaning and analyses were conducted using SAS 9.4 (Cary, NC). R (version 3.2.2) was used to calculate near-zero variance.

**Results**

*Participants*

Owners of 805 dogs expressed interest in the Golden Years study. Of these, 46 did not meet eligibility criteria (37 had history of or concurrent LPD other than TZL, 6 were non-cases <9 years of age, 3 were not purebred Golden Retrievers), 246 were lost to follow-up (at least 12 dogs died before enrollment was completed), and five owners only completed the demographic portion of the survey. In total, 508 dogs (140 cases, 221 subclinicals, 147 controls) were eligible and completed enrollment (both a blood sample and owner-completed questionnaire). These dogs represented at least 46 states (one unknown) and Canada (Figure 4.4). Descriptive statistics for all variables, stratified by disease status, are shown in Table 4.1. The median age of the case population was 10 years (range 7 - 15), compared to 11 years (range 9 - 18) for subclinicals and 10 years (range 9 - 15) for controls. Approximately 45% the population was male, independent of diseases status. The majority of our population received multiple preventive care measures, including flea and heartworm prevention (84% and 91% respectively) as well as regular rabies vaccinations (93%).

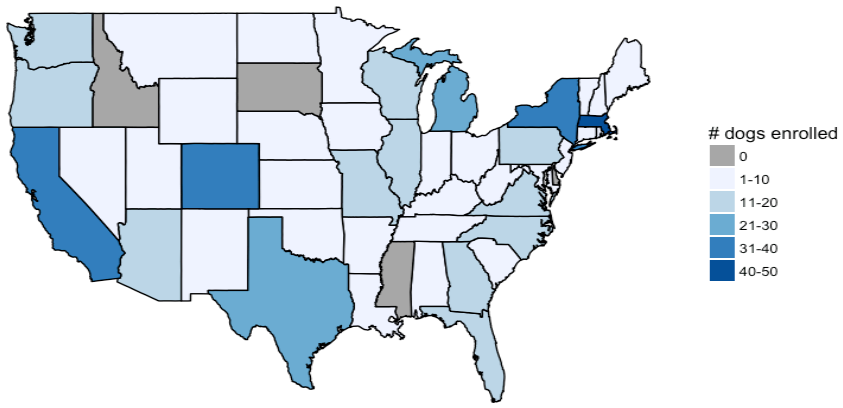



**Figure 4.4. State of residence for 508 dogs enrolled in the study.** Not depicted: one dog that resided in Alaska, two from Canada, one with unknown zip code of residence, and two with invalid zip codes.

*Univariable results*

Ten variables were significantly ($p<0.05$) associated with being an overt TZL case (vs. control) (Table 4.1). Variables univariably associated with an increased risk of TZL included history of or concurrent mange and vaccination frequency (receiving Distemper-Parvovirus combination (DHPP) vaccination as directed and receiving Bordetella vaccination "frequently"). Variables univariably associated with a decreased risk of TZL included age at spay/neuter, history of or concurrent hypothyroidism, receiving non-prescription supplements (both omega-3s and "other" supplements), frequent exposure to rural environments, and frequent swimming in irrigation water or lakes/streams. Two variables (bladder infection and eye disease) were significantly ($p<0.05$) associated with an increased risk of being subclinical (vs. control).

*Case versus control modeling*

Variables that met each criterion to be included in best subsets analysis are shown in Figure 4.5. Twenty variables were considered in best subsets analysis. We assessed collinearity for a model including all 20 variables; the highest condition index was 14.5, suggesting collinearity was not a problem in this model. Based on the results from best subsets regression, we divided variables into three "tiers" until we reached $\geq$14 degrees of freedom. The resulting 9-variable model (14df) was considered our "full" model. From there, our elimination strategy removed frequency of swimming in lakes/streams, resulting in an 8-variable model (Figure 4.5). Based on the likelihood ratio test, the full model did not perform significantly better than the reduced model ($p=0.15$). Three variables in this model were borderline significant (p-value between 0.05 and 0.06), but when the next variable (eye disease) was removed, there was

concern for confounding as the effect estimates for two variables (bladder infection and mange) changed by >10%.

**Table 4.1. Distribution of variables, stratified by disease status and recruitment strategy.**

| | | | | Cases | Controls | | | | Subclinicals | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | n missing | All (n=140) n (%) | All (n=147) n (%) | Database (n=111) n (%) | Clinic (n=36) n (%) | | All (n=221) n (%) | Database (n=164) n (%) | Clinic (n=57) n (%) | |
| Signalment | Age (years) | *median (range)* | 0 | 10 (7 - 15) | 10 (9 - 15) | 10 (9 - 15) | 11 (9 - 15) | | 11 (9 - 18) | 10 (9 - 18) | 11 (9 - 15) | |
| | Sex | Female (vs. male) | 0 | 76 (54%) | 85 (58%) | 67 (60%) | 18 (50%) | | 120 (54%) | 96 (59%) | 24 (42%) | * |
| | Age at spay/neuter | Female <1yr | 0 | 45 (32%) | 32 (22%) | 23 (21%) | 9 (25%) | * | 47 (21%) | 32 (20%) | 15 (26%) | * |
| | | Female >1yr | | 24 (17%) | 48 (33%) | 42 (38%) | 6 (17%) | | 68 (31%) | 61 (37%) | 7 (13%) | |
| | | Male < 1yr | | 35 (25%) | 37 (25%) | 26 (23%) | 11 (31%) | | 36 (16%) | 18 (11%) | 18 (32%) | |
| | | Male >1yr | | 22 (16%) | 23 (16%) | 18 (16%) | 5 (14%) | | 61 (28%) | 47 (29%) | 14 (25%) | |
| | | Missing | | 14 (10%) | 7 (5%) | 2 (2%) | 5 (14%) | | 9 (4%) | 6 (4%) | 3 (5%) | |
| Preventive care | Flea preventives | Ever (vs. never) | 9 | 116 (84%) | 121 (83%) | 92 (84%) | 29 (83%) | | 180 (83%) | 135 (84%) | 45 (82%) | |
| | Heartworm preventives | Ever (vs. never) | 17 | 124 (91%) | 130 (92%) | 100 (93%) | 30 (88%) | | 191 (90%) | 145 (91%) | 46 (87%) | |
| | Rabies vaccination | As directed (vs. not) | 4 | 128 (93%) | 137 (94%) | 104 (94%) | 33 (94%) | | 205 (93%) | 150 (91%) | 55 (96%) | |
| | DHPP vaccination | As directed (vs. not) | 8 | 119 (88%) | 100 (68%) | 70 (63%) | 30 (86%) | * | 145 (66%) | 100 (62%) | 45 (79%) | * |
| | Bordetella vaccination | Frequent (vs. not) | 0 | 80 (57%) | 70 (48%) | 52 (47%) | 18 (50%) | | 113 (51%) | 78 (48%) | 35 (61%) | |
| | | Missing | | 13 (9%) | 6 (4%) | 3 (3%) | 3 (8%) | | 10 (5%) | 6 (4%) | 4 (7%) | |
| | Teeth cleaned | Ever (vs. never) | 3 | 58 (42%) | 64 (44%) | 45 (41%) | 19 (53%) | | 98 (45%) | 70 (43%) | 28 (49%) | |
| Infectious diseases | Worms | Yes (vs. no) | 0 | 17 (12%) | 16 (11%) | 12 (11%) | 4 (11%) | | 28 (13%) | 24 (15%) | 4 (7%) | |
| | Bladder infection | Yes (vs. no) | 0 | 10 (7%) | 5 (3%) | 4 (4%) | 1 (3%) | | 30 (14%) | 23 (14%) | 7 (12%) | |
| | Mange | Yes (vs. no) | 0 | 16 (11%) | 3 (2%) | 1 (1%) | 2 (6%) | | 5 (2%) | 4 (2%) | 1 (2%) | |
| | Other infectious disease | Yes (vs. no) | 0 | 25 (18%) | 35 (24%) | 31 (28%) | 4 (11%) | * | 52 (24%) | 41 (25%) | 11 (19%) | |
| Inflammatory disorders | Eye disease | Yes (vs. no) | 1 | 21 (15%) | 14 (10%) | 10 (9%) | 4 (11%) | | 39 (18%) | 32 (20%) | 7 (12%) | |
| | Uveitis | Yes (vs. no) | 1 | 7 (5%) | 7 (5%) | 3 (3%) | 4 (11%) | | 10 (5%) | 7 (4%) | 3 (5%) | |
| | Ear Infection | Yes (vs. no) | 1 | 65 (46%) | 53 (36%) | 41 (37%) | 12 (34%) | | 102 (46%) | 73 (45%) | 29 (51%) | |
| | Tooth or Gum disease | Yes (vs. no) | 17 | 22 (16%) | 20 (14%) | 15 (14%) | 5 (14%) | | 27 (13%) | 21 (13%) | 6 (11%) | |
| | Gastrointestinal disease | Yes (vs. no) | 0 | 19 (14%) | 12 (8%) | 10 (9%) | 2 (6%) | | 27 (12%) | 18 (11%) | 9 (16%) | |
| | Cruciate ligament rupture | Yes (vs. no) | 0 | 8 (6%) | 7 (5%) | 5 (5%) | 2 (6%) | | 9 (4%) | 5 (3%) | 4 (7%) | |
| | Degenerative joint disease | Yes (vs. no) | 0 | 22 (16%) | 23 (16%) | 19 (17%) | 4 (11%) | | 41 (19%) | 31 (19%) | 10 (18%) | |
| | Non-mange skin disease | Yes (vs. no) | 0 | 60 (43%) | 53 (36%) | 38 (34%) | 15 (42%) | | 89 (44%) | 61 (37%) | 28 (49%) | |
| | Vaccine reaction | Ever (vs. never) | 2 | 4 (3%) | 12 (8%) | 9 (8%) | 3 (8%) | | 20 (9%) | 16 (10%) | 4 (7%) | |
| Other diseases | Hypothyroid | Yes (vs. no) | 0 | 8 (6%) | 24 (16%) | 22 (20%) | 2 (6%) | | 40 (18%) | 33 (20%) | 7 (12%) | |
| | Other endocrine disorder | Yes (vs. no) | 0 | 6 (4%) | 2 (1%) | 1 (1%) | 1 (3%) | | 3 (1%) | 2 (1%) | 1 (2%) | |
| | Cardiovascular disease | Yes (vs. no) | 0 | 6 (4%) | 10 (7%) | 8 (7%) | 2 (6%) | | 13 (6%) | 11 (7%) | 2 (4%) | |
| | Non-lymphoma cancer | Yes (vs. no) | 0 | 21 (15%) | 36 (24%) | 25 (23%) | 11 (31%) | | 55 (25%) | 33 (20%) | 22 (39%) | * |
| Medications | Antihistamines | Ever (vs. never) | 0 | 56 (40%) | 63 (43%) | 49 (44%) | 14 (39%) | | 106 (48%) | 81 (49%) | 25 (44%) | |
| | | Missing | | 38 (27%) | 35 (24%) | 25 (23%) | 10 (28%) | | 41 (19%) | 29 (18%) | 12 (21%) | |
| | Anti-inflammatories | Ever (vs. never) | 0 | 67 (48%) | 87 (59%) | 72 (65%) | 15 (42%) | * | 126 (57%) | 94 (57%) | 32 (56%) | |
| | | Missing | | 32 (23%) | 21 (14%) | 13 (12%) | 8 (22%) | | 35 (16%) | 24 (15%) | 11 (19%) | |
| | Steroids | Ever (vs. never) | 0 | 50 (36%) | 45 (31%) | 34 (31%) | 11 (31%) | | 87 (39%) | 69 (42%) | 18 (32%) | |
| | | Missing | | 40 (29%) | 38 (26%) | 28 (25%) | 10 (28%) | | 44 (20%) | 31 (19%) | 13 (23%) | |
| | Antibiotics | Ever (vs. never) | 0 | 103 (74%) | 118 (80%) | 94 (85%) | 24 (67%) | * | 173 (78%) | 130 (79%) | 43 (75%) | |
| | | Missing | | 25 (18%) | 14 (10%) | 6 (5%) | 8 (22%) | | 29 (13%) | 23 (14%) | 6 (11%) | |
| | Immunosuppressants | Ever (vs. never) | 0 | 4 (3%) | 8 (5%) | 7 (6%) | 1 (3%) | | 2 (1%) | 2 (1%) | 0 (0%) | |
| | | Missing | | 64 (46%) | 52 (35%) | 35 (32%) | 16 (44%) | | 84 (38%) | 58 (35%) | 26 (46%) | |
| | Omega-3s | Yes (vs. no) | 0 | 14 (10%) | 34 (23%) | 28 (25%) | 6 (17%) | | 66 (30%) | 52 (32%) | 14 (25%) | |
| | | Missing | | 30 (21%) | 28 (19%) | 21 (19%) | 7 (19%) | | 32 (14%) | 22 (13%) | 10 (18%) | |
| | Other Supplement | Yes (vs. no) | 0 | 38 (27%) | 67 (46%) | 55 (50%) | 12 (33%) | | 98 (44%) | 71 (43%) | 27 (47%) | |
| | | Missing | | 30 (21%) | 28 (19%) | 21 (19%) | 7 (19%) | | 32 (14%) | 22 (13%) | 10 (18%) | |
| Environmental exposures | Lawn exposure | Frequent (vs. not) | 6 | 56 (41%) | 72 (49%) | 57 (51%) | 15 (43%) | | 108 (50%) | 83 (52%) | 25 (44%) | |
| | Rural environment | Frequent (vs. not) | 0 | 46 (33%) | 72 (49%) | 59 (53%) | 13 (36%) | | 109 (49%) | 85 (52%) | 24 (42%) | |
| | | Missing | | 6 (4%) | 10 (7%) | 6 (5%) | 4 (11%) | | 8 (4%) | 5 (3%) | 3 (5%) | |
| | Parks | Frequent (vs. not) | 0 | 54 (39%) | 52 (35%) | 40 (36%) | 12 (33%) | | 54 (39%) | 33 (20%) | 21 (37%) | * |
| | | Missing | | 6 (4%) | 5 (3%) | 2 (2%) | 3 (8%) | | 12 (5%) | 8 (5%) | 4 (7%) | |
| | Tick on dog | Ever (vs. never) | 4 | 77 (55%) | 93 (64%) | 72 (65%) | 21 (58%) | | 133 (61%) | 98 (60%) | 35 (63%) | |
| Toxins | Lawn chemicals | Frequent (vs. not) | 10 | 17 (13%) | 9 (6%) | 4 (4%) | 5 (14%) | * | 20 (9%) | 15 (9%) | 5 (9%) | |
| | Cigarette smoke | Frequent (vs. not) | 16 | 3 (2%) | 10 (7%) | 9 (8%) | 1 (3%) | | 10 (5%) | 8 (5%) | 2 (4%) | |
| | Paints/solvents | Frequent (vs. not) | 19 | 6 (5%) | 3 (2%) | 2 (2%) | 1 (3%) | | 7 (3%) | 5 (3%) | 2 (4%) | |
| Swimming | Ocean | Frequent (vs. not) | 0 | 4 (3%) | 1 (1%) | 1 (1%) | 0 (0%) | | 6 (3%) | 4 (3%) | 2 (4%) | |
| | | Missing | | 35 (25%) | 44 (30%) | 35 (32%) | 9 (25%) | | 75 (34%) | 60 (37%) | 15 (26%) | |
| | Irrigation water | Frequent (vs. not) | 0 | 5 (4%) | 18 (12%) | 17 (15%) | 1 (3%) | * | 23 (10%) | 16 (10%) | 7 (12%) | |
| | | Missing | | 10 (7%) | 6 (4%) | 3 (3%) | 3 (8%) | | 19 (9%) | 12 (7%) | 7 (12%) | |
| | Lakes/streams | Frequent (vs. not) | 0 | 12 (9%) | 27 (19%) | 24 (22%) | 3 (8%) | * | 29 (13%) | 22 (13%) | 7 (12%) | |
| | | Missing | | 6 (4%) | 8 (5%) | 6 (5%) | 2 (6%) | | 10 (5%) | 6 (4%) | 4 (7%) | |

*p<0.05 when comparing database-recruited to clinic-recruited by Fisher's exact test

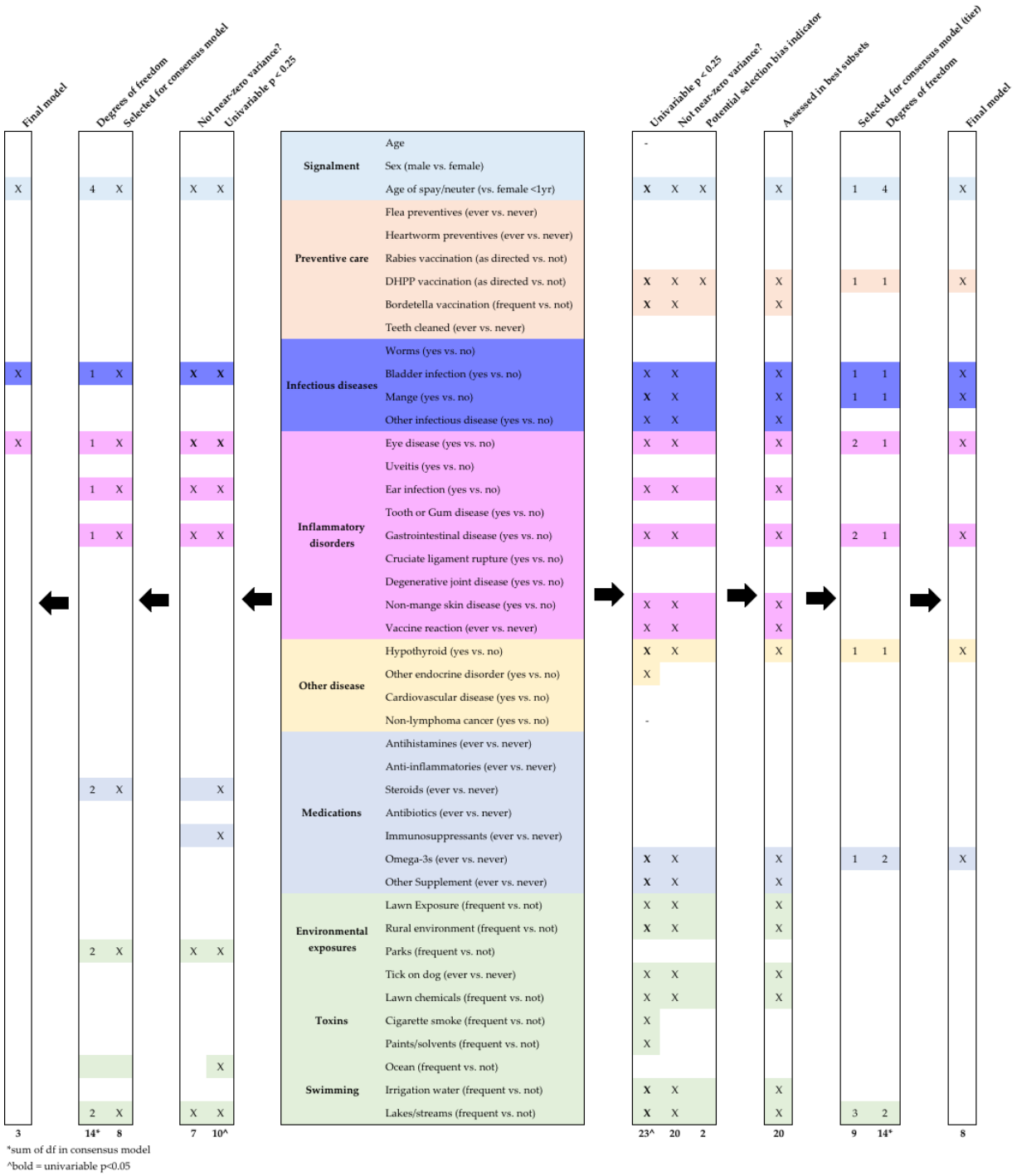



**Figure 4.5. Flow chart of variables selected based on criteria to be assessed in multivariable modeling, best-subsets, and backward selection.**

Our final model indicated four inflammatory/infectious diseases were associated with increased odds of TZL (Table 4.1; Figure 4.6). These diseases included history of or concurrent mange (OR 5.5, 95% CI 1.4 – 21.1), bladder infection (OR 3.5, 95% CI 0.96 – 12.7), gastrointestinal disease (OR 2.4, 95% CI 0.98 – 5.8), and eye disease (OR 2.3, 95% CI 0.97 – 5.2). These were generally uncommon outcomes, with mange present in only 11% of TZL cases and 2% of controls, bladder infection present in 7% of TZL cases and 3% of controls, gastrointestinal disease present in 14% of cases and 8% of controls, and eye disease present in 15% of cases and 10% of controls. In addition, three variables were associated with significantly decreased odds of TZL, including hypothyroidism (OR 0.25, 95% CI 0.10 – 0.66), receiving omega-3 supplements (OR 0.29, 95% CI 0.13 – 0.63), and being a female spayed at >1 year of age (vs. <1 year) (OR 0.34, 95% CI 0.16 – 0.74). DHPP vaccination frequency was also significantly associated with increased odds of TZL (OR 3.2, 95% CI 1.6 – 6.5).

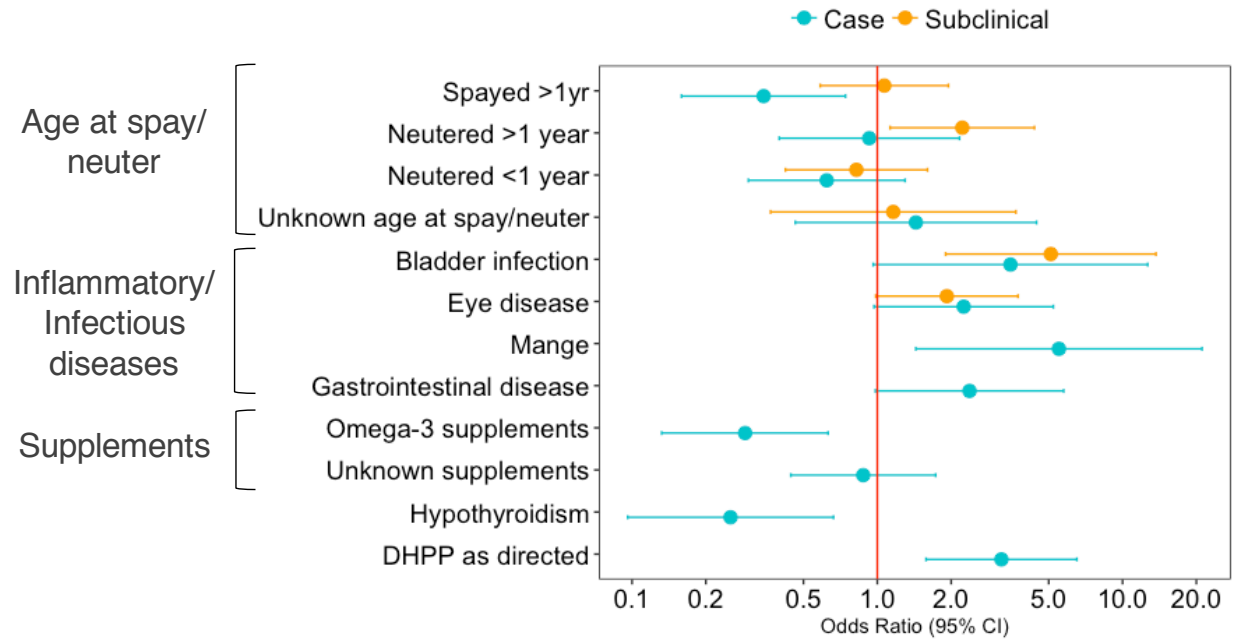



**Figure 4.6. Final multivariable modeling results for case vs. control model and subclinical vs. control model.** ORs and 95% CIs are shown in log scale.

*Subclinical versus control*

Eight variables met the criteria to be considered in best subsets analysis (Figure 4.5). Since these variables accounted for 14 degrees of freedom and our sample size allowed for 14, we did not proceed with best subsets and instead included all variables in the initial full model. When we assessed collinearity for this model, the highest condition index was 7.2, suggesting collinearity was not present. Our elimination strategy removed all variables except bladder infection, eye disease, and age at spay/neuter (Figures 4.5 and 4.6). Eye disease was borderline significant (OR 1.9, 95% CI 0.99 – 3.75), but the likelihood ratio test suggested this model performed better than the model without eye disease (p=0.048). Unlike the case vs. control model, the difference in age at spay/neuter for subclinicals vs. controls appeared to be driven by males neutered at >1 year of age (vs. females spayed at <1 year) (OR 2.2, 95% CI 1.1 – 4.4). In addition, dogs with history of or concurrent bladder infections had a five times greater odds of being subclinical (vs. control) (OR 5.1, 95% CI 1.9 – 13.7). Bladder infection and eye disease were relatively uncommon in subclinicals, present in 14% and 18%, respectively (vs. 3% and 10% of controls).

*Systematic differences in recruitment strategy*

For most variables, the proportion of dogs did not differ based on recruitment strategy (Table 4.1). Only two variables differed significantly among both controls and subclinicals: age at spay/neuter and DHPP vaccination. Differences in age at spay/neuter appeared to be largely driven by a higher proportion of female database-recruited dogs being spayed at >1 year of age (37% among database-recruited subclinicals vs. 13% and 17% among clinic-recruited [p=0.01]).

However, information on timing of spay/neuter was more likely to be missing among clinic-recruited controls (14% vs. 2% among database-recruited [p=0.04]). For DHPP vaccination frequency, database-recruited dogs appeared less likely to be vaccinated as directed (62% and 63% among database-recruited subclinicals and controls vs. 79% and 86% among clinic-recruited [p=0.02 for both comparisons]). Upon further investigation, this appeared to be driven by an increased likelihood of these owners to use titers to dictate vaccination as opposed to a conventional vaccination schedule.

Both DHPP vaccination frequency and age at spay/neuter were considered potential markers of selection bias in our multivariable models for case vs. control. Once a final model was chosen, we removed these variables one-by-one to determine if confounding existed. When DHPP vaccination frequency was removed, ORs for spaying at >1 year of age, bladder infection, hypothyroidism, and gastrointestinal disease changed by >10%. When age at spay/neuter was removed, ORs for DHPP vaccination frequency, mange, and hypothyroidism changed by >10%. Likelihood ratio tests suggested our chosen final model performed better than the models without DHPP vaccination frequency (p=0.001) or age at spay/neuter (p=0.024).

*Shared risk factors for cases and subclinicals*

History of or concurrent bladder infection and eye disease were associated with increased odds for being either a case or a subclinical (Figure 4.6). The individual ORs for cases and subclinicals differed slightly, but the confidence intervals had substantial overlap. Eye disease was a heterogeneous category comprised of cataracts, glaucoma, progressive retinal atrophy or degeneration, and sudden-onset blindness. These individual outcomes were rare (1-

3% of participants) except cataracts, which were present in 14% of cases, 15% of subclinicals, and 9% of controls.

Mange and gastrointestinal disease were only included in the multivariable model for cases and did not appear to be important predictors for subclinical vs. control analyses. Gastrointestinal disease was univariably associated with subclinicals (p=0.22), but did not remain significant in the multivariable setting. This was also a heterogeneous group of diseases, including colitis, diarrhea, gastritis, and "other" gastrointestinal diseases. These outcomes were generally evenly distributed, with each present in 1-5% of participants.

Age at spay/neuter was also significant for both models, but the category driving the association differed. Being spayed at >1 year of age (vs. spayed <1 year of age) was associated with decreased odds of being a case, whereas being neutered at >1 year of age was associated with increased odds of being subclinical. Omega-3 supplements, hypothyroidism, and DHPP vaccination frequency were only associated with risk among cases.

*Sensitivity analyses*

We removed 20 cases <9 years of age and saw no appreciable differences in effect estimates or variable significance (Suppl. Table 4.1). Removal of 112 dogs with non-lymphoma cancer (21 cases, 55 subclinicals, 36 controls) led to model instability, evidenced by large confidence intervals (Suppl. Tables 4.1 and 4.2). However, effect estimates followed the same trends.

**Discussion**

The etiology of specific subtypes of PTCL remains unclear, but chronic antigen stimulation is suspected to play a role. We undertook a large case-subclinical-control study of TZL among Golden Retrievers to investigate the hypothesis that chronic inflammatory conditions increase TZL risk. Our study found multiple inflammatory risk factors with diverse mechanistic pathways were associated with an increased risk of TZL. Consistent with a prior report,[28] we found a significant association of mange and TZL. Additionally, in our multivariable models, bladder infections and eye disease had borderline associations with both TZL cases and subclinicals. The power to detect an association for bladder infections and eye disease was low due to small sample sizes, but the consistency of these associations for both cases and subclinicals gives us some confidence that these are biologically important associations. We also noted a similar borderline increased risk with gastrointestinal disease and TZL. Dogs who were administered omega-3 supplements, which may reduce inflammation,[171, 177-179] had a decreased risk of TZL. Hypothyroidism, which can arise through autoimmune mechanisms,[161, 180] was also associated with a decreased risk of TZL. Consistent with prior studies,[61, 172] late spay (>1 year of age) was associated with a decreased risk of TZL. In contrast, late neuter was associated with an increased risk of being subclinical.

The role of a chronic antigen stimulus in the development of PTCL has been suggested by both laboratory and epidemiologic studies.[5, 63, 103] It is believed that persistent stimulation of T cells causes sustained proliferation, allowing accumulation of mutations that lead to lymphoma. Prior studies have attempted to pinpoint a specific antigen or disease process responsible for

lymphoma development, but have had limited success. We believe this suggests TZL may be driven by persistent immune stimulation as a general process as opposed to a specific antigen.

*Variables associated with increased TZL risk*

Consistent with a prior report,[28] we found 11% of cases (n=16) had a recent or concurrent diagnosis of mange, compared to three control dogs (2%) and five subclinical dogs (2%). Because of the often concurrent diagnosis of mange and TZL, it is unclear whether mange is a risk factor for TZL or whether TZL causes immunosuppression, making dogs susceptible to mange. Unfortunately, we could not fully evaluate temporality in our dataset as we had wide age categories that generally overlapped with age at diagnosis. Prospective studies would be beneficial in elucidating the natural history of TZL and mange.

Bladder infections, more generally urinary tract infections (UTIs), are most commonly bacterial in origin.[181, 182] Bacterial UTIs affect 14% of all dogs during their lifetime, making them the most common infectious disease of dogs.[182, 183] However, veterinary patients are often asymptomatic and thus may go undiagnosed; studies have shown asymptomatic bacteriuria may be present in 2-14% of dogs.[184, 185] This may also suggest that while UTIs are often an acute finding, they may represent underlying chronic inflammation. To our knowledge, only one study[186] has evaluated UTIs (categorized as cystitis, prostatitis and pyelonephritis) as a risk factor for human non-Hodgkin's lymphoma (NHL), and found increased risk of prostatitis for overall NHL, but not T-cell NHL. This may be a novel association for TZL, or perhaps may reflect differences in UTI etiology for people and dogs. While UTIs are most common among

spayed female dogs,[187] we believe bladder infection is an independent risk factor since our multivariable models adjust for sex and neuter status.

Eye disease and gastrointestinal disease are crude categories that incorporate multiple disease mechanisms. Gastrointestinal diseases were relatively evenly split across the specific disease categories (colitis, diarrhea, gastritis). These categories are nonspecific and often interrelated, so reliability of owner report may be poor. In general, gastrointestinal disease may be acute or chronic and can thus represent many etiologies. Diarrhea, for example, is a nonspecific diagnosis and can occur through numerous mechanisms, including infectious (e.g. parasites, viruses, bacteria), toxic (e.g. pesticides, plants), inflammatory (e.g. inflammatory bowel disease), and diet-related etiologies (e.g. indiscriminate eating, adverse reaction to food). The chronic forms of gastrointestinal disease are typically inflammatory in origin, and thus may play a more important role in TZL pathogenesis. However, multiple bouts of acute gastrointestinal disease could also be a source of persistent immune perturbation. Prior studies[186, 188] have found no association of ulcerative colitis, inflammatory bowel disease, or gastroenteritis with overall human NHL, but the association may have been masked by using overall NHL instead of specific T cell subtypes. Certain long-term gastrointestinal infections, including *Helicobacter pylori* and *Campylobacter jejuni* have been associated with human mucosa-associated lymphoid tissue lymphomas,[189-191] highlighting the possibility that these are subtype-specific risk factors.

With regard to eye disease, association with TZL appears to be driven predominately by cataracts in both cases and subclinicals. In dogs, cataracts are characterized by a vision-altering opacity in the lens or its capsule that develops due to disruption of the lens fibers.[192, 193]

Cataracts are most commonly an inherited disease, but can also form via trauma or secondary to diabetes mellitus.[192, 193] Lenticular sclerosis, which is caused by a normal increase in nuclear density as dogs age, also causes a cloudy appearance to the eye[192-194] and thus may be misreported as cataracts by owners. As this study relies on owner-reported diagnoses, we are unable to differentiate true cataracts from lenticular sclerosis. Both diseases are common in older dogs (>9 years)[192-194] and the mechanism by which they would influence TZL risk is unclear. To our knowledge, no prior studies have assessed eye diseases as a risk factor for lymphoma.

*Variables associated with decreased TZL risk*

Use of omega-3 supplements was associated with a decreased risk of TZL. Omega-3 supplementation may have anti-inflammatory and immunomodulatory effects,[171, 177-179] so it is possible that their supplementation can mediate the effects of inflammatory diseases. This finding is in agreement with previous studies that suggest omega-3 supplementation is associated with a decreased risk of human lymphoma[195, 196] and can delay onset of T cell lymphomas in mice.[197] In future studies, it would be valuable to ascertain detailed information about omega-3 use, such as dose and duration of use. This would also aid in the ability to evaluate effect modification of omega-3 use and inflammatory disorders.

In dogs, hypothyroidism is generally caused by destruction of the thyroid gland by either lymphocytic thyroiditis (likely immune-mediated) or idiopathic atrophy.[161] Dogs with hypothyroidism were significantly less likely to have TZL, which is in contrast to previous studies that have found autoimmune disorders increase the risk of multiple lymphoma

subtypes, including PTCL[63, 64, 87, 188, 198] However, the positive metabolic and growth-promoting effects of thyroid hormone offer a plausible explanation for why this particular form of autoimmune disease might be protective. Individuals with Hashimoto's thyroiditis, a type of human hypothyroidism, can develop primary thyroid lymphoma, although this outcome is rare.[199] One study found a positive association of Hashimoto's thyroiditis and overall human NHL, which is in contrast to our findings.[198] In general, dogs are only tested for hypothyroidism if there is clinical suspicion, so dogs who were considered hypothyroid-free in our dataset may have never been tested. We were not able to adjudicate hypothyroidism diagnosis in our dataset. However, this is likely to be less of an issue in our study as this is a common disease in Golden Retrievers and our owners are prudent about veterinary care, evidenced by high use of preventive care measures (flea and heartworm prevention). In a medical record review of 50 hypothyroid-free dogs in our dataset, 12 (24%) had been tested and were negative. Overall, the potential mechanism through which hypothyroidism may reduce TZL risk is still unclear, and may require future laboratory studies.

*Variables that may be indicators of selection bias*

One concern with this study was potential selection bias resulting from the database-recruited dogs. This population represents owners who are highly invested in Golden Retriever health, many of whom show and/or breed Golden Retrievers. It is possible that this population is not representative of the case source population with respect to exposures of interest, thereby raising the concern of selection bias.

We addressed this concern by conducting stratified analyses for all variables, and suspected selection bias for variables with significant differences between recruitment strategies (database vs. clinic) among both subclinicals and controls. Based on this criteria, DHPP vaccination frequency and age at spay/neuter remain potential indicators of selection bias, but we believe controlling for these variables reduced bias in our effect estimates (Figure 4.3).

A potential role of sex hormones is plausible based on prior research.[61, 172] Among female dogs, spaying after 1 year of age resulted in a significantly decreased risk of TZL. This finding is supported by recent evidence that late spay/neuter may reduce incidence of many diseases, including lymphoma (non-significant reduction for females, significant reduction for males).[61] The finding that neutering after 1 year resulted in a significantly increased risk of being subclinical is in contrast with prior lymphoma research; however, no studies have assessed risk among subclinicals, which may reflect a pre-neoplastic state. While we expect male and female sex hormones to act via different mechanisms, it is difficult to explain why they would affect cases and subclinicals differently. One possible explanation is that this is a spurious finding due to differences in comparison-group recruitment strategies, as anecdotal evidence suggests database-recruited dogs were more likely to be breeding bitches who were spayed late or kept intact. This would explain why late spay appeared associated in the case vs. control comparison, but not in the subclinical vs. control comparison.

An association with vaccination status was not one we considered *a priori*. We suspect DHPP vaccination frequency represents a selection bias in our control recruitment, as database-recruited controls had a higher use of titers, and thus a decreased vaccination frequency. This may also explain why only DHPP vaccination frequency was associated with TZL, as opposed

to rabies vaccination frequency, which is dictated by state laws, and Bordetella vaccination frequency, which is not a core vaccine and thus depends on the dog's lifestyle and likelihood of exposure.

Subclinical vs. control status was not known at the time of enrollment and survey completion. Since the proportion of subclinicals and controls was equal among both recruitment strategies (60% of database-recruited and 61% of clinic-recruited dogs were subclinical), we hypothesize that any bias in recruitment strategy should affect both subclinicals and controls equally, leading to unbiased effect estimates in subclinical vs. control models. Furthermore, an indicator variable for recruitment strategy was not significantly associated with subclinical vs. control status (data not shown; chi-square p=0.81), suggesting there were no systematic differences by recruitment strategy for this comparison.

*Factors not associated with TZL*

There were many variables that fit our chronic inflammation hypothesis that were not found to be associated in our study, including infectious diseases, skin disease, and environmental exposures. These variables were significantly associated with TZL in the univariable setting, but not in multivariable models, which could suggest they were highly correlated with other variables or were poorly measured. Alternatively, it is possible these variables do not play a role in TZL pathogenesis. Prior research has found inconsistent evidence for associations of eczema, atopic dermatitis, hay fever, and allergy with human NHL and PTCL,[63, 200, 201] so it is unclear whether they have an etiologic role. A prior study found living or working on a farm was inversely associated with human PTCL.[103] It is possible that the hygiene

hypothesis plays a role in this, dampening Th2 cell reactivity and thus decreasing the chance of developing a lymphoproliferative disorder. From the human cancer literature, one study indicated that as public hygiene conditions improve, childhood acute lymphoblastic leukemia incidence increases.[202] Future investigations into these variables may be warranted as we cannot confidently differentiate measurement error from a lack of biologic significance in our study.

*Strengths and limitations*

Prior epidemiologic studies of PTCL have been limited by small sample sizes, forcing researchers to consider PTCL as a whole instead of evaluating subtype-specific risk factors. Our study considers only one subtype of PTCL among a relatively genetically homogeneous population, so we have a greater ability to detect risk factors. In addition, we use a comprehensive questionnaire to evaluate disease history and environmental exposures, allowing us to assess a greater scope of risk factors than prior studies. To our knowledge, this is the largest study assessing subtype-specific risk factors for canine lymphoma.

Another strength of our study is the ability to rule out disease in the control population, minimizing misclassification of disease. It is often not possible to rule out cancer in the control population due to lack of screening tests or limited funding. However, we were able to test all controls for the presence of TZL cells, thereby ruling out any undiagnosed cases. It is still possible our control dogs will develop TZL later in life, but we have minimized that likelihood by selecting older dogs and ruling out current disease. Because subclinicals have not previously been described, there is no clear diagnostic criteria for this group of dogs. We chose a systematic

flow cytometric cutpoint to minimize the number of subclinicals in the control group and thereby maximize specificity.

Measurement error is a concern with self-reported questionnaires, especially because we asked owners to recall exposures and disease histories over the course of their dog's lifetime. Since the questionnaire was long and comprehensive and owners did not know our exposure of interest, we believe it is unlikely that there was extreme over- (or under-) reporting of our exposures of interest. However, measurement error likely arose through the collapsing of variables due to 1) sparseness of data and 2) our hypothesis that owners were unable to distinguish between certain diseases. In both these circumstances, collapsing of variables led to an inability to evaluate distinct disease mechanisms individually, which likely attenuated effect estimates. It also restricted consideration of age at diagnosis for heath history variables to incorporate a lag time to diagnosis of TZL. Future studies may be able to obtain larger sample sizes and more detailed information, including veterinarian-confirmed diagnoses, for the variables suggested to be important in our study.

*Conclusions*

Overall, this study contributes to the body of evidence that chronic immune stimulation may be important in TZL pathogenesis. The ability to identify subclinicals poses a unique opportunity to evaluate factors involved in progression from subclinical to case, and prospective follow-up studies are warranted. Additionally, longitudinal studies such as the Golden Retriever Lifetime Study,[32] will be paramount in furthering our understanding of the

natural history of TZL, through obtaining detailed health and exposure histories over the course of a dog's life.

# CHAPTER 5: CONCLUSIONS

My overall goal of this dissertation was to investigate environmental and genetic risk factors for mature, antigen-driven lymphoproliferative disorders using a canine model. However, my wider-reaching goal was to contribute to the framework of canine cancer models and begin to elucidate novel ways we can use dogs to further our understanding of both mechanistic and epidemiologic aspects of human disease. As such, I will divide my conclusions for this dissertation into these two parts. First, I will discuss what my research means in the context of canine and human lymphoma research, specifically among mature, antigen-driven subtypes. I will summarize key findings from each aim, questions that remain unanswered, and discuss how they might inform future research. Second, I will discuss how this work furthers the canine model. I will review important lessons I have learned through conducting this research and considerations for future research using the canine model.

**Conclusions in the context of lymphoma research**

Despite the substantial burden of lymphoma, the etiology of most subtypes remains unclear. Two indolent subtypes, B-CLL and TZL, are believed to arise from mature, antigen-driven lymphocytes. Current research suggests both genetic and environmental factors play a role in the etiology of these diseases. By studying these diseases in parallel, we can increase our understanding of potential shared mechanisms underlying the development of indolent, mature, antigen-driven lymphoproliferative disorders. We addressed this goal through three

aims, under the hypothesis that human and canine lymphoproliferative disorders share risk factors, clinical presentation, and prognosis.

- **Aim 1**: Describe canine B-CLL presentation, clinicopathologic findings, and breed predisposition
- **Aim 2**: Identify genes associated with TZL in Golden Retrievers using a genome-wide association study
- **Aim 3**: Examine the association of chronic inflammatory conditions and TZL using a questionnaire-based case-control study

*Aim 1: B-cell chronic lymphocytic leukemia*

To establish the canine B-CLL model, we had three main objectives. First, we assessed whether the clinical signs and laboratory findings are similar for canine and human B-CLL. Second, we evaluated a potential breed predisposition for canine B-CLL. Third, we determined whether any breeds have a unique B-CLL presentation. Below, I provide a brief rationale for each objective, summarize key findings, and outline future directions.

Aim 1 Objective 1: Evaluate B-CLL as a model for human B-CLL

While we know canine B-CLL is cytologically equivalent to human B-CLL, little is known about the actual clinical presentation. A first key step in evaluating canine B-CLL as a model for human B-CLL is determining whether the disease presents similarly in patients with regard to clinical signs and laboratory findings. To do this, we summarized physical examination and clinicopathologic findings of 491 dogs at the time of B-CLL diagnosis. From this analysis, we identified many similarities between canine and human B-CLL. For both, this

is typically a disease of older individuals, with the average age at diagnosis 70 years in people[203] and 11 years in dogs. We saw that about half of dogs presented with lymphadenopathy and splenomegaly, and about 30% with hepatomegaly. These clinical signs correspond to the human Rai Stages I and II and are associated with a more favorable prognosis.[109] Only about a quarter of dogs presented with anemia, and less than 10% with thrombocytopenia, which are seen in Rai Stages III and IV and confer a worse prognosis. As B-CLL is often an incidental diagnosis,[203] we expect most dogs to be diagnosed at an early stage of disease as suggested by these findings. The main difference we saw between canine and human B-CLL is that human B-CLL is about twice as common in males vs. females,[204, 205] whereas we saw no sex difference in canine B-CLL. We suspect this finding is due to spaying and neutering within our canine populations, but further research needs to be conducted to evaluate how sex hormones influence B-CLL risk.

Overall, these findings indicate canine B-CLL has many similarities with human B-CLL, suggesting it may be an appropriate model for future studies. A next key step in evaluating the canine B-CLL model is to conduct follow-up cohort studies to gain a greater understanding of prognostic factors and the natural history of canine B-CLL. Specifically, it would be informative to determine whether clinical signs at presentation correlate with the same staging (and thus outcome) as human B-CLL. We are currently conducting a follow-up study of a subset of dogs included in this study to help address this knowledge gap.

Another next step in establishing the canine model is to evaluate whether canine and human B-CLL share the same gene expression profile. Our laboratory is addressing this through two studies. First, we conducted a NanoString (NanoString Technologies, Inc., Seattle, WA) experiment using gene probes for genes known to be up- or down-regulated in human B-CLL.

Preliminary data suggests at least 70% of these genes are also differentially expressed in the same direction in canine B-CLL. Since this study design only allows evaluation of pre-selected genes, we will also conduct an RNA-sequencing experiment for a more agnostic evaluation of differentially expressed genes. This will allow us to identify any gene expression profiles that are unique to canine B-CLL.

Aim 1 Objective 2: Identify whether there is a breed predisposition for canine B-CLL

In people, first-degree relatives of B-CLL patients have an 8.5-fold higher risk for developing the disease.[10] However, a clear set of risk-associated genes has not been identified. Since breeds of dogs represent genetically similar individuals, we hypothesized that genetic risk for canine B-CLL would be evidenced by a breed predisposition. We found a significantly increased odds of B-CLL among 10 small breed dogs (Bichon Frise, Boston Terrier, Cairn Terrier, Cocker Spaniel, Dachshund, Jack Russell Terrier, Maltese, Pomeranian, Shih Tzu, and Yorkshire Terrier) and two larger breeds (English Bulldog and Pit Bull).

By conducting genetic studies within one predisposed breed of dog, we may have an enhanced ability to identify genetic risk factors. Our laboratory is in the early stages of planning a GWAS in Shih Tzus for these purposes, and we have demonstrated the feasibility and utility of conducting such studies in aim 2. The increased risk of B-CLL among small breed dogs suggests the genetic risk factor may be related to breed size. It has been demonstrated that almost half the variation in dog size can be explained by a few genetic variants.[117, 118] It is possible that one of these genes is mechanistically involved in B-CLL pathogenesis. Alternatively, genes associated with B-CLL may be near genes involved in dog size, causing

them to often be inherited together. Results from a GWAS study will help elucidate potential genes associated with B-CLL risk.

Aim 1 Objective 3: Evaluate breed-specific patient presentation

As researchers have evaluated human B-CLL, it has become clear that there are discrete prognostic subtypes that may reflect distinct risk factors. We were interested in whether specific dog breeds may have unique presentations of B-CLL, potentially signifying genetically distinct subtypes. As such, we compared patient presentation for the top six breeds with significantly increased odds of B-CLL. Compared to mixed breed dogs, English Bulldogs presented at a significantly younger age (6 years vs. 11 years). In addition, they had decreased expression of class II MHC and CD25, which may indicate an immature phenotype.[114, 115] It is possible other breeds have unique presentations of B-CLL, but additional analysis was not performed due to limited sample size.

If English Bulldogs have an immature phenotype of B-CLL, we would expect a shortened survival time. In human B-CLL, $IgV_H$ mutational status is an important predictor of overall survival, with unmutated $IgV_H$ genes conferring substantially reduced survival.[110] Work in our laboratory has indicated that Boxers with B-CLL are more likely to have unmutated $IgV_H$ genes. Thus, a key goal of our follow-up study is to compare B-CLL survival for English Bulldogs, Boxers, and "typical" B-CLL (represented by small breed dogs). We suspect survival will be variable, as with human B-CLL, because a prior study found survival ranged from 30 to 1,000 days.[88] Results from this study will further establish the canine B-CLL model by showing that dogs have similar prognostic factors to humans.

For the NanoString and RNA sequencing experiments, we are also evaluating gene expression differences for English Bulldogs and Boxers vs. "typical" B-CLL. Preliminary principal components analysis of NanoString data shows these three groups form distinct clusters, suggesting unique gene expression profiles. Genes upregulated in Boxers vs. English Bulldogs were involved in cell proliferation and cell cycle, potentially indicating a more active or aggressive phenotype. The RNA sequencing experiment will validate these differences and may also identify unique genes that distinguish B-CLL subtypes. These experiments will provide insights into the pathogenesis of B-CLL and how gene expression relates to phenotype and prognosis.

If evidence continues to support Boxers having an aggressive, unmutated B-CLL subtype, we will conduct a GWAS within this breed. This could provide a unique opportunity to evaluate heritable risk for an aggressive B-CLL subtype that has been difficult to evaluate in human populations, in part due to their genetic diversity.

Aim 1 Summary

Since B-CLL is relatively common in people, its pathogenesis, treatment, and prognosis is better understood than TZL. However, as researchers have investigated B-CLL, they have identified prognostic subtypes that may reflect distinct risk factors. By establishing the canine B-CLL model and determining breed predisposition, we can target breeds for future analyses, enhancing our ability to identify genes of interest. In addition, identifying breed-specific subtypes can enable easier identification of study subjects and increased power for genetic and epidemiologic analyses. Findings from these studies are also valuable for veterinary medicine,

as it is important to identify dogs who will have a poor prognosis so we can treat them appropriately.

*Aims 2 and 3: T zone lymphoma*

Research on the pathogenesis of human B cell lymphomas has been crucial in developing improved diagnostics and treatments.[5] However, our understanding of human T cell lymphomagenesis is lagging behind, largely due to the heterogeneity of T cell lymphomas and low incidence of individual subtypes. Due to small sample sizes, epidemiologic studies have been forced to evaluate PTCL as a whole, instead of investigating individual subtypes independently. This leads to biased results and an inability to differentiate risk factors for each subtype. Since T cell lymphomas are comparatively more common in dogs (35% of lymphomas vs 9% in humans),[21] we can attain larger sample sizes in our canine populations than we can with people. Thus, using the canine model is an important first step toward reaching adequate sample sizes to assess PTCL subtype-specific risk factors.

Aims 2 and 3 investigated genetic and environmental risk factors for one PTCL subtype, TZL, among a genetically homogeneous population (Golden Retrievers), thereby increasing our ability to identify key etiologic factors. *We hypothesized that a genetic risk factor predisposes dogs to TZL, but chronic inflammation is necessary for progression to clinical disease* (Figure 5.1). The study population used for these aims was a case-subclinical-control study of 508 Golden Retrievers aged 9 years and older. All owners completed a questionnaire for use in the aim 3 risk factor analysis; a subpopulation of 360 dogs were genotyped for the GWAS in aim 2. Conclusions for

these aims are divided into 1) findings regarding genetic and environmental risk for TZL, and 2) how our findings contribute to our understanding of subclinical dogs.

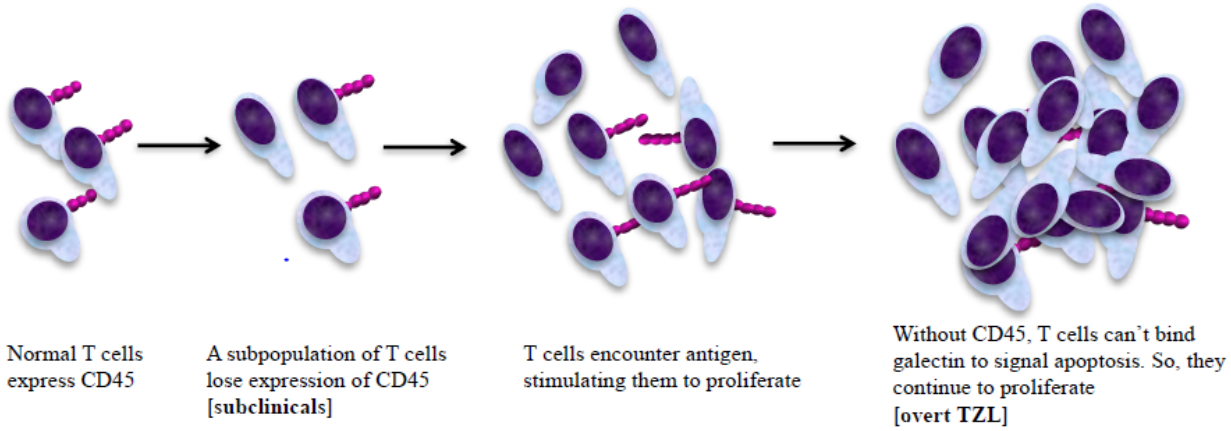



**Figure 5.1. Hypothesized origin of TZL.**

Aims 2 and 3 Objective 1: Evaluate genetic and environmental risk factors for TZL

*Objective 1a: GWAS (aim 2) findings*

Over 40% of TZL cases occur in Golden Retrievers,[6] suggesting a genetic component of this disease. Using a GWAS, we identified two main association signals, one on chromosome 8 and the other on chromosome 14. Targeted resequencing identified five coding mutations within three hyaluronidase genes clustered on chromosome 14, one of which was predicted to be possibly damaging. We did not identify a clear coding mutation on the chromosome 8 peak, but potential pathways implicated include vesicle trafficking; alterations in this pathway could impact antigen presentation or lead to accumulation of misfolded proteins. Additionally, we identified mutations that could influence thyroid hormone regulation.

Our chromosome 14 findings indicate a potential role of hyaluronidases in TZL pathogenesis. Hyaluronan breakdown has been associated with pro-inflammatory and pro-oncogenic byproducts, so the association with TZL is biologically plausible. This finding is

particularly interesting because the coding mutations we identified were shared with a recent paper of mast cell tumors in Golden Retrievers.[128] Follow-up laboratory work is essential to better understand the functional consequence of these mutations and thereby how they may influence disease risk. Using immunohistochemistry, Arendt et al.[128] found increased hyaluronan staining in mast cell tumor samples compared to normal skin and pannicular fat. Likewise, it would be informative to measure both hyaluronidase and hyaluronan levels in tumor vs. normal tissue of TZL cases.

We conducted a NanoString experiment to assess expression of SPAM1 and HYAL4 among a separate population of TZL cases and controls (not limited to Golden Retrievers) and preliminary analysis suggests no significant difference. However, there may be downstream changes not captured in our NanoString experiment. RNA sequencing may be useful to identify other differentially expressed genes, particularly those that may be markers of an inflammatory response. We are also interested in quantifying high vs. low molecular weight hyaluronan in TZL cases and controls, as increased levels of high molecular weight hyaluronan have been deemed important in cancer resistance among naked mole rats.[140]

Our chromosome 8 targeted resequencing did not implicate one particular gene, but suggested pathways which may be important. Our NanoString experiment also assessed expression of SEL1L, STON2, CEP128, and TSHR. Preliminary analysis suggests TZL cases have significantly decreased SEL1L expression and significantly increased STON2 expression. Decreased SEL1L expression could lead to a reduced ability to breakdown misfolded proteins, ultimately causing endoplasmic reticulum stress and an unfolded protein response.[206, 207] Future work could involve measuring biomarkers of this pathway in both TZL cases and controls to

assess their role in TZL pathogenesis. Increased STON2 expression could alter antigen trafficking and expression by antigen presenting cells.[156] This could subsequently alter T cell activation, promoting lymphomagenesis. Measuring biomarkers to T cell activation could provide insight into this potential mechanism, although prospective analysis would be necessary to differentiate whether T cells are activated as a consequence of neoplastic transformation or whether activation played a role in initiating neoplastic transformation. As numerous pathways are plausible based on our chromosome 8 findings, a full RNA sequencing experiment may help distinguish which pathways are dysregulated.

This GWAS study provided insight into potential mechanisms underlying TZL pathogenesis. Future work can build upon our results, further investigating these mechanisms and better understanding the pathophysiology leading to disease. On a larger scale, the ability to narrow down specific pathways involved in TZL pathogenesis could ultimately inform development of therapeutics to better treat this disease.

*Objective 1b: Questionnaire (aim 3) findings*

We hypothesized that chronic antigen stimulation is important in TZL pathogenesis. Thus, we conducted a questionnaire-based analysis evaluating diseases and exposures that could cause chronic inflammation. We found multiple inflammatory risk factors, including bladder infection, mange, eye disease, and gastrointestinal disease increase TZL risk. In addition, use of omega-3 supplements,[171, 177-179] which reduce inflammation and have immunomodulatory effects, reduced TZL risk. We also found hypothyroidism, generally believed to be an autoimmune disease,[161, 180] reduced TZL risk.

Since our findings were largely owner-reported health-related risk factors, a prudent next step would be adjudicating these diagnoses with medical records. This would give us more detailed diagnostic information and allow us to group diseases by potential mechanistic pathway for further analysis. Additionally, we could better evaluate timing of associated disease diagnosis, particularly in the context of timing of TZL diagnosis. There also may be a critical window of exposure to environmental toxicants that was unable to be evaluated in our study. A prospective study, such as the Golden Retriever Lifetime Study, could provide insight into timing and dose of exposures as they relate to TZL development. Prospective studies also enable more efficient data ascertainment as they are not subject to recall bias or missing data due to inability to obtain medical records retrospectively. Additionally, results from NanoString and RNA sequencing experiments may help elucidate the potential mechanisms by which associated inflammatory conditions may influence TZL risk.

We were unable to determine whether the observed associations with DHPP and age at spay/neuter were true associations or due to a systematic bias from control recruitment strategies. Future studies may be able to differentiate this by using a control series that better represents the source population of cases. Our clinic-recruited control series likely would have provided unbiased effect estimates, but our sample size was too limited to only use this population. It is worth noting, however, that there was no difference in DHPP vaccination frequency or percent of females spayed at >1 year of age between cases (88% received DHPP as directed, 17% spayed >1year) and clinic-recruited controls (86% received DHPP as directed, 17% spayed >1 year).

With respect to veterinary medicine, results from aim 3 could be used to identify dogs who may be at a higher risk of TZL. Our findings suggest omega-3 supplementation could help reduce TZL risk, however more research is needed to replicate this finding and to determine the required dose. We are particularly interested in, and are currently in the process of, following dogs with TZL over time to determine overall survival and natural history of disease. Prior studies have indicated median survival is long (>2 years) regardless of treatment,[6, 28] so we hypothesize cases ultimately die of non-TZL-related causes. However, preliminary owner-reported data suggests a handful of cases have been euthanized due to chronic infections. It is possible TZL leads to immunosuppression, leading to chronic, unresolving infections. We hope our large follow-up study will help inform veterinarians and owners about what to expect when a dog is diagnosed with TZL.

*Objective 1c: Tying together GWAS (aim 2) and questionnaire (aim 3) findings*

Both aims 2 and 3 implicated a potential role of inflammation in TZL pathogenesis, in line with our original hypothesis (Figure 5.1). Multiple inflammatory risk factors, including bladder infection, mange, eye disease, and gastrointestinal disease were found to be associated with TZL in aim 3. Aim 2 also implicated potential inflammatory pathways on both chromosome 8 and 14. Alterations in hyaluronidase genes could cause inflammation through increased hyaluronan breakdown.[140, 141] Additionally, hyaluronan can directly stimulate the inflammasome. Altered regulation of SEL1L and STON2 can cause endoplasmic reticulum stress, which can increase transcription of pro-inflammatory cytokines and additionally activate the inflammasome.[206, 207] It is plausible that inflammation due to genetic changes could predispose development of the inflammatory diseases we found to be associated with TZL in

aim 3, and those may thus either be intermediates in the path to TZL development (Figure 5.2a) or simply descendants of the same ancestor (Figure 5.2b). Prospective studies would help differentiate between these alternatives. The protective association of omega-3 supplementations observed in aim 3 is also in line with an important role of inflammation, as omega-3 supplements reduce inflammation.[171, 177-179] Future work should investigate the interaction of omega-3s with identified genetic and environmental factors to determine whether it has potential as a therapeutic.

While TZL and mast cell tumors share variants in hyaluronidase genes, we did not see an increased risk of mast cell tumors among dogs with TZL. This may suggest that hyaluronidase mutations predispose dogs to cancer, but different environmental stimuli influence which cancer the dog will ultimately develop. Conducting a mast cell tumor case-control questionnaire analysis like that in aim 3 could help identify factors that may drive dogs to develop mast cell tumor vs. TZL.

Alterations in T cell activation through changes in vesicle trafficking could influence development of many other diseases. Development of autoreactivity, for example, could lead to hypothyroidism or inflammatory bowel disease.[161, 180, 208] Additionally, decreased T cell function could predispose dogs to infectious diseases, such as bladder infection or mange. The same potential interplay of variables discussed with inflammation is thus possible with T cell activation, where health disorders may either be intermediates on the causal path (Figure 5.2c) or descendants of the same ancestor (Figure 5.2d). Omega-3 supplementation could likewise influence these pathways through its immunomodulatory effects.[177-179] Biologic interaction could also play a role, where the combination of altered T cell activation and inflammatory disorders

leads to increased risk of TZL development. A greater understanding of both biologic mechanisms and the natural history of TZL is necessary to tease apart these relationships.

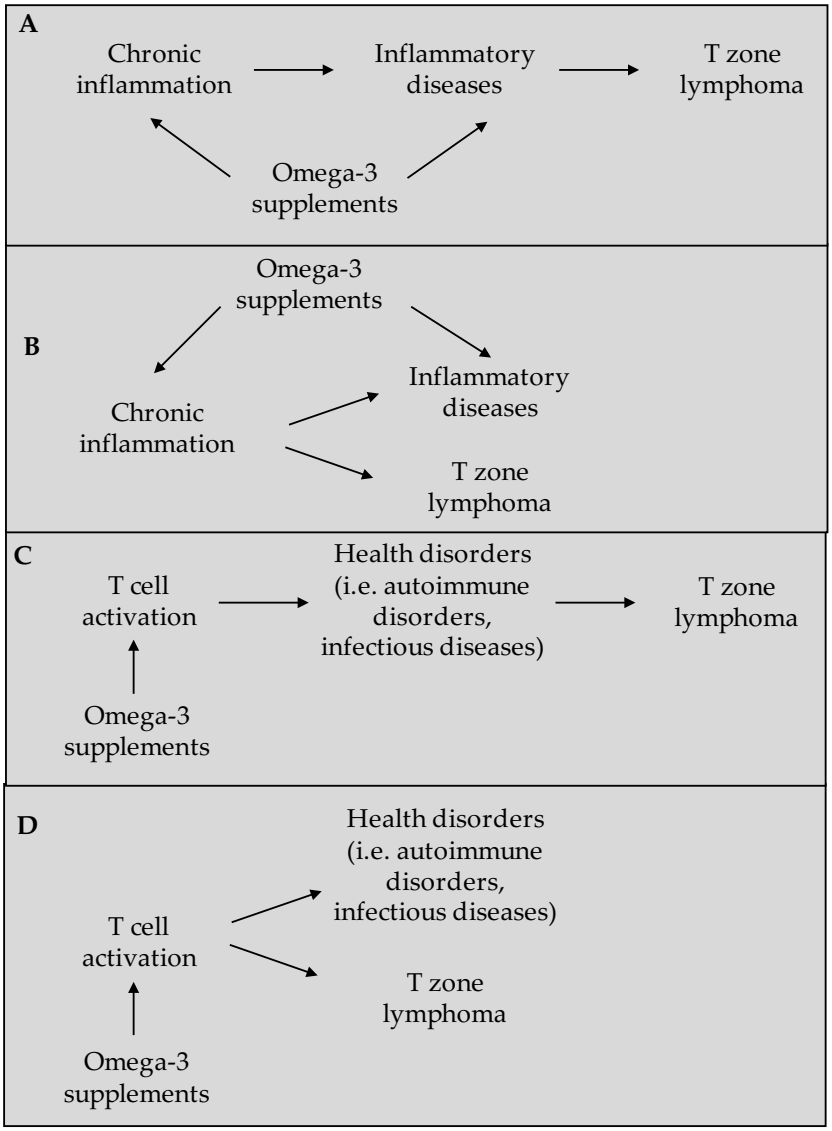



**Figure 5.2. Directed acyclic graphs (DAGs) explaining potential relationship between aim 2 and 3 findings.** (A) Genetic changes lead to chronic inflammation, which predispose dogs to inflammatory diseases and subsequently TZL; omega-3 supplements can reduce inflammatory responses. (B) Genetic changes lead to chronic inflammation, which predisposes dogs to both inflammatory diseases and TZL; no causal relationship exists between inflammatory diseases and TZL. (C) Genetic changes lead to alterations in T cell activation, which predisposes dogs to autoimmune disorders and/or infectious diseases; these diseases play a role in TZL pathogenesis. Omega-3 supplements can have immunomodulatory effects, altering T cell activation. (D) Genetic changes lead to alterations in T cell activation, which predisposes dogs to both health disorders and TZL; no causal path exits between the health disorders and TZL.

Lastly, the parallel finding of a hypothyroidism association in aim 3 and a GWAS association near genes involved in thyroid hormone regulation (DIO2 and TSHR) is worth noting. While we did not find any coding mutations in these genes, there is the potential for regulatory changes that are not yet understood. We conducted a LiftOver of this region to the human genome, and determined some mutations were in regions predicted to be regulatory for human DIO2. Unfortunately, the LiftOver for the TSHR region was poor and most SNPs did not map to the human genome, hindering our ability to evaluate TSHR regulatory elements. Thyroid hormone is important in many functions throughout the body, and thus levels are intricately managed.[162, 209] Activity of deiodinase 2, the product of DIO2, is known to be regulated by endoplasmic reticulum stress.[164] As such changes in SEL1L or STON2 could have downstream impacts on levels of T3, the active form of thyroid hormone.

As thyroid hormone plays an important role in cell growth and metabolism, it is possible that lack of this hormone may decrease T cell proliferation and therefore help prevent the development of TZL. It would be informative to conduct *in vitro* studies to assess the influence of thyroid hormone levels on T cell proliferation and development. We could also assess T cell function at the time of hypothyroidism diagnosis (i.e. before treatment with exogenous thyroid hormone) compared to normal dogs to determine whether there are changes in T cell activation. It is alternatively possible that the observed association is due to hypothyroidism therapeutics as opposed to the disease itself. Pharmaceutical researchers should evaluate whether there are differences between exogenous and endogenous thyroid hormone which may confer protection against TZL.

Overall, aims 2 and 3 support the hypothesis that chronic inflammation has an etiologic role in TZL. More research is necessary to validate and elucidate this finding. If these findings are replicated, it may indicate that veterinarians should more closely monitor dogs with signs of chronic inflammation. Additionally, research on potential therapeutic benefits of anti-inflammatory medications may be warranted.

*Aims 2 and 3 Objective 1 Summary*

Findings from aims 2 and 3 may help elucidate potential mechanisms behind TZL lymphomagenesis, guiding future research to improve diagnostics and treatments. Ideally, we would confirm our findings in a human TZL population to demonstrate this research is translatable. However, TZL is incredibly rare in people,[98] so it may be difficult to establish whether our canine TZL findings are translatable to human TZL. Nevertheless, these studies demonstrated that it is feasible to conduct epidemiological studies in a specific subtype of T cell lymphoma, and thus may provide a broader impact in the understanding of T cell lymphomas. In particular, we could build upon this work by applying a similar study design to the more aggressive PTCL subtype seen in dogs. This subtype is likely more reflective of the aggressive PTCL that afflicts people and thus may be a more impactful model to understand human lymphoma.

Aims 2 and 3 Objective 2: Where do subclinicals fit into the picture?

When we began this study, we knew we could identify small amounts of $CD45^-$ T cells in the blood of Golden Retrievers, but we were unsure how common they were among aged Golden Retrievers. This led to our hypothesis that the presence of $CD45^-$ T cells was likely dictated by a genetic risk factor and that progression to clinically apparent TZL required an

environmental stimulus, such as chronic antigen stimulation (Figure 5.1). Due to the case-control design, we were not able to directly evaluate the timing of these factors with respect to progression in this study. However, our results can provide mechanistic insights for future studies. We predicted a genetic risk factor was responsible for the presence of TZL cells, and thus that cases and subclinical cases would share the same genetic risk factors in aim 2. We then predicted that a chronic antigen stimulus was required for these cells to progress to clinical disease, and thus that in aim 3 inflammatory disorders would appear as a risk factor among cases, but not among subclinical cases. Despite these predictions, we used a data-driven approach in our analyses to agnostically determine whether to consider subclinicals independently or combine them with either cases or controls.

Among human lymphoproliferative disorders, only B cell lymphoma precursors have been identified, including monoclonal gammopathy of undetermined significance and monoclonal B cell lymphocytosis (MBL).[210, 211] The identification of a TZL precursor opens many doors in research, from following patients prospectively to understand how and why the disease progresses to developing screening tools to intervene early in the disease process. This dissertation provides the basis to conduct such studies in T cell lymphomas.

*Objective 2a: GWAS-related findings (aim 2)*

We believe subclinicals represent a pre-neoplastic state, so this state must precede development of overt TZL. This is seen in MBL, where MBL always precedes B-CLL,[212] although overall risk of transformation is low (1-2% per year).[213] SNPs associated with B-CLL are also associated with MBL,[214] congruent with the belief that they are two stages of the same disease. Similarly, we predicted a genetic risk factor was responsible for the presence of TZL cells. As

such, we expected cases and subclinicals would share the same genetic risk factors. However, we found no significant differences between subclinicals and controls in our GWAS analysis. Thus, our findings may represent genes associated with susceptibility for developing overt TZL as opposed to risk factors for developing TZL cells.

There are three possible explanations for our inability to differentiate subclinicals and controls in our study. First, it is possible that the genetic risk factor for developing $CD45^-$ T cells is shared among all Golden Retrievers, making it impossible to identify within our dataset. In the canine genome, it is possible to get sections of DNA that are completely homozygous within a breed, which is evidenced by the number of SNPs with a fixed minor allele frequency. Future work could involve conducting pathway analysis for genes within the regions of high homozygosity in Golden Retrievers to determine if any pathways are biologically plausible for the presence of $CD45^-$ T cells. Additionally, conducting a GWAS in another breed of dogs that commonly develop TZL, such as Shih Tzus, may help identify a shared region of homozygosity. If Shih Tzus do not have a shared region of homozygosity, we may be able to identify a significantly associated peak in Shih Tzus that corresponds to a homozygous region in Golden Retrievers.

Second, it is possible that the associated genetic differences, while present, may not be identified within a GWAS study. The evaluation of heritable risk for a disease is complex, and a GWAS cannot capture the full complexity. Whole genome sequencing or whole exome sequencing may be warranted to better identify the associated region. Additionally, it is possible that multiple genes interact to confer risk of developing $CD45^-$ T cells. If this is the case,

novel statistical methods and increased sample sizes may be needed to identify important genes.

Third, it is possible that a biological process, perhaps due to an environmental exposure that the majority of Golden Retrievers encounter, causes downregulation of CD45 on T cells and only dogs with the genetic risk factors identified in our study are susceptible to TZL progression. This seems like the least plausible explanation for our findings, as it is difficult to hypothesize an environmental factor that would be shared by such a large, dispersed population.

*Objective 2b: Survey-related findings (aim 3)*

For aim 3 analyses, we chose to analyze subclinicals as a separate group because we identified statistically significant differences when compared to cases and controls. We found bladder infection, eye disease, and being a male neutered at >1 year of age were all associated with an increased risk of being subclinical. Eye disease and bladder infection were also risk factors for TZL. While we cannot evaluate temporality from our analyses, we hypothesize that variables associated with TZL but not subclinicals might be factors involved in progression. Likewise, it is possible that omega-3 supplementation, which was associated with a decreased risk of TZL, may be used to prevent progression among dogs with $CD45^-$ T cells.

Key next steps involve following subclinicals prospectively to understand the natural history of disease. This project is currently underway in our laboratory, and will help answer fundamental questions including the overall risk of progression, average time to progression, and clinical signs associated with progression. This information is essential to determine the health relevance and burden of subclinicals. From a veterinary perspective, this information can

better inform clinicians and clients what to expect if their dog has subclinical CD45$^-$ T cell levels. In addition, it can be used to develop guidelines for frequency of re-evaluation among these dogs.

Follow-up of our subclinical population can also be used to identify risk factors for TZL progression. As part of our follow-up study, we are obtaining owner-reported health updates and terminal medical records for use in these purposes. This will help us establish temporality of specific risk factors, particularly mange. Understanding risk factors for TZL progression is an important first step in reducing the burden of this disease. If the risk factors we identify are preventable, we could help decrease the occurrence of progression by instructing owners to avoid certain exposures. Non-preventable risk factors are also valuable because they can help us identify dogs who may need to be monitored more closely or potentially treated sooner.

Additional laboratory-based studies of interest include evaluation of cytokine levels, particularly Th2-related cytokines, at various time points from initial identification of CD45$^-$ T cells to TZL progression. This may help elucidate subtle inflammatory changes involved in progression that may not be clinically apparent. In addition, it would be informative to conduct RNA sequencing experiments comparing CD45$^-$ T cells in subclinicals vs. cases. Differences in gene expression may elucidate important changes necessary for progression and could ultimately be used to help identify dogs at higher risk of progressing to overt TZL.

*Other subclinical considerations*

Our laboratory conducted a study to determine whether the presence of CD45$^-$ T cells is unique to Golden Retrievers or simply a common finding in older dogs (Hughes et al., manuscript under review). We discovered these cells are significantly more common in Golden

Retrievers than other breeds. However, we did not evaluate other breeds that commonly develop TZL, such as Shih Tzus. It would be informative to conduct a study to determine whether all breeds that commonly get TZL have a high prevalence of subclinicals, or whether Golden Retrievers are unique. This information could also inform expected results for a GWAS in Shih Tzus; if the prevalence of subclinicals is much lower, we may expect to find a separate associated peak instead of a shared homozygous region.

Additionally, determining the age dogs tend to develop $CD45^-$ T cells and whether there are risk factors for the initial appearance of these cells would help elucidate TZL etiology. A prospective study such as the Golden Retriever Lifetime Study could be used to sequentially test for the presence $CD45^-$ T cells and determine the average age at appearance. The questionnaire and medical data they obtain could subsequently be used to evaluate risk factors for the presence of $CD45^-$ T cells.

*Aims 2 and 3 Objective 2 Summary*

A clearer understanding of subclinicals could inform future development of biomarkers to identify dogs that should have more intensive clinical supervision. In addition, gene expression findings could be used to develop more effective treatments, either to target genes important in malignant transformation and progression, or to better target malignant cells in overt cases.

Aims 2 and 3 Summary

While findings from these studies cannot directly evaluate TZL pathogenesis, they provide insights for future studies. We identified risk factors unique to cases (mange, gastrointestinal disease, genetic variants on chromosomes 8 and 14) that may reflect risk for

progression from subclinical to case. Additionally, we found that hypothyroidism and omega-3 supplementation may prevent that progression. We also identified factors that may be associated with risk of developing CD45$^-$ T cells, including eye disease and bladder infection. We suspect there are additional genetic factors associated with risk of developing CD45$^-$ T cells that we were not able to identify in our GWAS study. These findings are summarized in Figure 5.3 below.

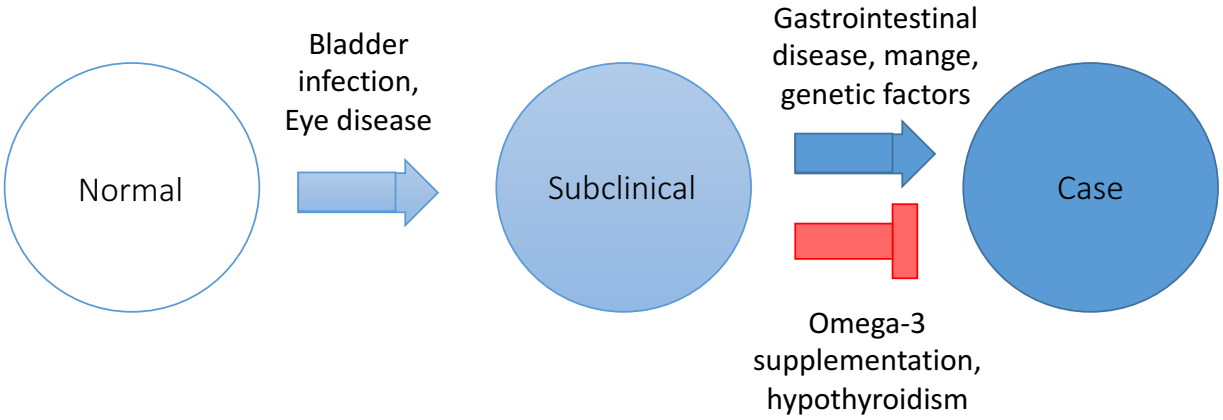



**Figure 5.3. Potential interpretation of results in the context of TZL pathogenesis.**

**The canine cancer model**

While the canine lymphoma model has been accepted for over a decade, the potential for epidemiologic impacts has largely been untapped. In this dissertation, we utilized canine populations to demonstrate novel uses of the canine model. In particular, we have demonstrated the feasibility of conducting observational epidemiologic studies through both national laboratories and nationwide case-control study recruitment. In this section, I will review important considerations learned through conducting these studies and how they can inform future canine epidemiologic studies. Specifically, I will highlight:

- Feasibility, strengths, and limitations of using national laboratories to conduct surveillance-type cancer studies (aim 1)
- Considerations when conducting canine GWAS and thoughts about the future of canine genetic studies (aim 2)
- Feasibility of conducting survey-based observational epidemiology in canine populations (aim 3)
- Strengths and limitations of clinic and database recruitment strategies for case-control studies (aim 3)

*Aim 1*

Conducting large-scale surveillance-type cancer studies in dogs is logistically more difficult than in humans. In people, there are many programs that work to gather data on an ongoing basis for the purposes of cancer surveillance. Programs such as the Surveillance, Epidemiology, and End Results Program (SEER; https://seer.cancer.gov/) provide publically available population-based data on demographics, tumor characteristics (i.e. grade, stage), and survival. SEER also links with U.S. mortality and population data to calculate incidence and compare trends over time. In addition, the National Cancer Database (https://www.facs.org/quality-programs/cancer/ncdb) is a hospital-based cancer registry in which data are collected by trained registrars. This database provides detailed tumor, treatment, and demographic information and is believed to represent more than 70% of newly diagnosed cancer cases in the U.S. In veterinary medicine, we do not have a systematic way to capture new cases nor do we have an ongoing census to provide a denominator for incidence or prevalence

calculations. Attempts have been made to provide similar infrastructure for canine populations, but have largely been unsuccessful.[47] The Veterinary Medical Database is the longest-running registry and contains data on >500,000 patients since 1996 (https://vmdb.org/). Unfortunately, currently very few hospitals submit data, so the representativeness is limited.

We do, however, have large national laboratories, such as IDEXX (https://www.idexx.com/small-animal-health/index.html) and Antech (http://www.antechdiagnostics.com), that obtain canine tumor samples from throughout the U.S. For studies that have a descriptive epidemiologic basis, such as our aim 1 analysis, these laboratories may provide the most comprehensive study population. The Clinical Immunology Laboratory database is similarly comprehensive, although the scope is smaller in that only lymphoproliferative diseases are represented. One of the major benefits of the Clinical Immunology Laboratory is the ability to readily obtain data on patient presentation for a large population of dogs with one subtype of lymphoma. This population can also form the basis for prognostic studies, either prospectively or retrospectively.

One of the most important considerations when conducting an analysis like our breed comparison in aim 1 is the choice of a reference group. Most diagnostic laboratories only obtain data from sick animals, so the typical case-control comparison where controls are free of disease is not valid. By nature of the contingency table used when calculating an odds ratio, we have two comparisons: 1) the reference disease, and 2) the reference breed. When using the Clinical Immunology Laboratory database, our choice of a reference disease is limited, as it must be another lymphoproliferative disease. Evidence suggests certain lymphoma subtypes have a breed predisposition, so care must be taken when choosing a reference subtype as this can skew

results. We chose all non-CLL LPDs as our comparison group, as we suspect the effect of subtype-specific breed predispositions may be attenuated by combining all subtypes. Recent data from our laboratory shows this may be the case as the overall proportion of breeds in our database closely reflects American Kennel Club breed registration. Thus, our calculated odds of B-CLL (vs. non-CLL LPD) within each breed may be a relatively unbiased estimate of the overall odds of B-CLL.

The second comparison choice is the reference breed. Ideally, we would compare to the overall risk of B-CLL across all breeds. However, since we do not have a defined cohort, the assumptions to calculate a risk ratio are not met. Instead, we have to choose a particular breed as a comparison. There are advantages and limitations to each choice, and weighing them depends on the study goals. Since our main interest was breed predisposition as a surrogate for genetic risk, we believed mixed breed dogs were an appropriate choice, as they may represent the genetic heterogeneity of dogs overall. In addition, they had a similar proportion of B-CLL as the overall population in our database. However, without genetic data we cannot know how well mixed breed dogs actually represent the genetic heterogeneity of dogs, and their representativeness could vary from study to study. By chance, we could have a skewed population of mixed breed dogs that ends up biasing our reference group toward a particular breed. To have a more homogeneous comparison population, we also looked at Labrador Retrievers. However, since Labrador Retrievers had a lower proportion of B-CLL, the resulting odds ratios were correspondingly skewed. As there is no clear "best" reference group, there is utility in conducting a sensitivity analysis to evaluate how choice of reference group influences effect estimates.

While national laboratories do not provide the same level of information as SEER or the National Cancer Database, they are a great resource for veterinary medicine. If care is taken to interpret results in the context of the comparison being made, much can be learned about potential differences in the distribution of disease across breeds. With relatively little time and cost, we can use diagnostic laboratories to gather a large population of dogs with reasonably standardized diagnostic criteria. Such studies could also be used to evaluate trends over time to identify any changes in the distribution of subtypes.

*Aim 2*

Hundreds of GWAS studies have been conducted in dogs. However, the body of methodological literature is far behind what is available for human genomic studies. While many considerations for human and canine GWAS are shared, information about the differences discussed in the background section of this dissertation, particularly accounting for relatedness and stratification both within and between breeds, is not readily available. While aim 2 taught me a great deal about conducting GWAS in Golden Retrievers, it is likely that different breeds have their own considerations. For instance, English Bulldogs have far fewer polymorphic loci than Golden Retrievers (86k vs 114k)[11] and thus it is possible fewer English Bulldogs are needed for a GWAS. However, residual population stratification may be a bigger issue in this breed due to a higher level of inbreeding and fewer founder sires. Thus, nuances in GWAS study design and analysis may differ for each breed. For any future canine GWAS studies it would be prudent to obtain pedigrees from all dogs to avoid selecting overly related

dogs for genotyping. This will prevent unnecessary costs of genotyping dogs who will later be removed with relatedness filters.

As sequencing costs have declined over time, more researchers are moving toward next-generation sequencing methods, such as whole exome sequencing and whole genome sequencing, to identify variants associated with disease. GWAS was designed with the "common disease-common variant" hypothesis. That is, the belief that common variants (SNPs) underlie the genetic basis of common diseases. Because of the way SNP chips are designed, it is unlikely that we will identify causative SNPs. Instead, these SNPs are "markers" for an actual associated SNP and follow-up work is necessary to identify causative mutations. Identification of the causative SNP requires next-generation sequencing of areas in linkage disequilibrium with the associated SNPs. This allows researchers to determine the exact location of variants and whether they are in coding or non-coding regions to better evaluate genes involved in disease pathogenesis.

In whole exome sequencing, only areas which code for functional proteins (exons; about 1-2% of the genome) are sequenced. Some researchers argue that 80-90% of inherited disease-causing mutations in Mendelian disorders are located in exons.[215, 216] Compared to GWAS, whole exome sequencing is beneficial in that we have an enhanced ability to identify rare variants. However, there is an increased cost and computational burden.

Whole genome sequencing has many benefits over whole exome sequencing, but with an even higher cost and computational burden. With whole genome sequencing, we gain more information than whole exome sequencing, including both coding and non-coding DNA. Non-coding DNA was originally thought to be "junk" DNA that served no purpose; however,

research has shown many of these areas have functional roles, particularly in regulation of gene expression. Whole genome sequencing allows us to evaluate variants in these areas, as well as in exons, which may be important in disease pathogenesis. Additionally, it allows the evaluation of other types of variants, including copy number variation, insertions, deletions, and inversions. Since whole genome sequencing does not require an enrichment step, there is more uniform coverage of the genome and thus less bias in read depth.

The increased cost of whole genome sequencing is not directly proportional to the percent of the genome sequenced, however, because whole exome sequencing requires more laboratory preparation (PCR and enrichment) and a higher sequencing depth (100x vs 30x). Overall, the cost of whole genome sequencing is about 3-4 times that of whole exome sequencing. Researchers must also consider the cost of computing storage space and additional analytical burden. The computational burden can be substantial, and since less is known about non-coding regions of DNA, more follow-up laboratory work may be necessary to understand mechanisms of pathogenesis. As technology improves and sequencing costs decrease, more researchers will likely move toward whole genome sequencing in order to maximize the amount of information obtained.

While next-generation sequencing technologies provide many benefits over GWAS, they are not without limitations: sequencing and subsequent alignment to genomes are error-prone and limited by the quality of the reference genome. Variants identified must be prioritized based on quality and likelihood of being "real". Thus, it is possible that important variants will go undetected. Computer scientists are working to improve algorithms to reduce these problems.[217, 218] Choice of genotyping vs. sequencing depends on many considerations, from

practicalities such as availability of data storage, laboratory capabilities, and available funds (which is a balance of number of dogs studied and cost per dog) to the research question you seek to answer and the analytic complexity you want to undertake.

As genetic analyses have increased in popularity, it has become increasingly apparent that many genetic variants only account for a small percent of disease susceptibility. One potential explanation for this is that genes do not act in isolation; they have complex interactions with other genes and with the environment. Regardless of the technology used, we have not yet developed a systematic way to evaluate gene-gene and gene-environment interactions. When we evaluate interactions, we quickly end up with sparse data, as the number of genotype combinations go up exponentially as each SNP is added to the model. In addition, it is computationally intensive to parse through all possible options when we have thousands of candidate SNPs, and this approach is prone to false positives. New algorithms are being developed to evaluate gene-gene interactions, including random forests and multifactor dimensionality reduction.[219-224] However, many of these algorithms require an initial SNP screening that often means only SNPs with marginal effects are evaluated further. These algorithms are still in early development, and have generally only been tested on simulated data or limited datasets. It is unclear whether they are robust when 1) the effect size of each SNP is minimal and 2) environmental factors or other covariates play a role in disease pathogenesis.

In general, there is a symbiotic relationship between laboratory-based research and sequencing-based research. We need to have a better understanding of gene function and regulation in order to interpret the biologic relevance of sequencing data. However, sequencing data is continually generating new findings that need to be explored in the laboratory.

Ultimately, an increased understanding of how the genome and biology relate will allow us to take a less agnostic approach to our analyses and incorporate what we know about pathobiology into our analytics. These scientific advances are tightly tied to improvements in statistical approaches, computing capabilities, and sequencing technologies. As technology continues to provide larger amounts of higher quality data, both faster and at a reduced cost, we must continue developing more efficient algorithms and modeling strategies that can accommodate numbers of variables that far exceed sample sizes. The simultaneous integration and advancement of these fields are integral to future successes.

Overall, the infrastructure for conducting canine genomic studies has vastly improved since the first canine genome was released in 2004. For GWAS, the same tools can be used for human and canine analyses. As the future is moving toward next-generation sequencing, tools for these analyses need to be optimized for canine analyses. Fortunately, the Broad Institute and Uppsala University are paving the way with their cutting-edge work. As more dogs are sequenced, annotations are improving and better references are becoming available. This work will increase our ability to draw meaningful conclusions from canine genomics studies.

*Aim 3*

Prior canine observational epidemiology has involved varied exposure assessment and recruitment strategies. Our work involved two novel approaches: an internet-implemented owner-completed questionnaire and two recruitment strategies.

Survey design and approach

We found the online survey resulted in high completion rates (78%) with few technical problems. The ability to automate emails and reminders in SurveyMonkey® saved time and allowed for simpler organization. The data was easy to download and only required minor manipulation for input into R or SAS. Due to the size of this study, it was necessary to use the upgraded version of SurveyMonkey®, which allows for unlimited questions and responses and a more customizable survey experience (i.e. question skip logic, custom themes). We only had two participants without access to a computer and/or internet connection: one completed the survey on a hard copy and faxed their responses, the other's veterinarian completed the survey online on their behalf. Overall, there were very few disadvantages to the use of the online questionnaire. It is becoming increasingly less common for people to answer phone calls, particularly from unknown phone numbers, so we believe online recruitment increases response rate. Additionally, the ability to conduct most communication via email and online questionnaire was invaluable for a fulltime veterinary student.

We sought to design a questionnaire that was relatively quick and easy for owners to complete, but was still comprehensive. With this goal in mind, there are a few modifications that would improve the questionnaire for future use. First, I would alter health history questions to incorporate all potential ages at diagnosis (rounded to nearest year) as opposed to preset age categories. While age groups may ultimately need to be collapsed, this would allow greater flexibility in doing so, particularly for sensitivity analyses assessing diagnoses prior to a certain age. Second, I would expand the list of diseases within each disease category. We originally tried to limit these lists to keep the survey short, but in hindsight most dogs did not

have substantial disease history and owners probably knew what disease they were looking for based on the initial category grouping (i.e. eye disease, endocrine disorder, etc.). Thus, this likely would only have a minor impact on survey completion time. An extended disease list would be beneficial for several reasons, including decreasing the amount of free text to parse through (in the "other" category) and a better ability to subsequently group diseases based on shared mechanisms. Ideally, we would also be able to obtain medical records to validate survey findings.

Participant recruitment

Clinic recruitment of cases was fairly successful and not remarkably time-intensive. We identified cases through submissions to the Clinical Immunology laboratory and contacted veterinarians to inform them of our study. Successful enrollment required cooperation from the veterinarian and willingness to contact the owner, obtain a consent form on a subsequent visit, and submit another blood sample (if necessary). Case owners and veterinarians were generally invested in helping to understand this disease; of dogs who were eligible for our study, about 75% ultimately consented and completed the survey. While we do not have comprehensive data on why the other 25% did not enroll, it appears a large portion were diagnosed through an oncology clinic and then returned to their referring veterinarian for subsequent care. Thus, the submitting veterinarian did not see the patient again to discuss enrollment with owners. This number could be improved in future studies by contacting referring veterinarians about enrollment in these circumstances.

The clinic-recruited control population was less successful than for cases and required more contact from our laboratory. Many veterinarians either were unable to find controls or did

not respond to our email requests. Among the 140 cases enrolled in the study, only 73 had controls recruited from the same clinic (total 135 control dogs), and 10 of the control dogs did not meet enrollment criteria (<9 years or had another lymphoproliferative disorder). However, enrollment was generally successful among the dogs that were recruited and met enrollment criteria, with 93% completing enrollment. It is understandable that veterinarians have busy schedules and recruiting dogs for our study is not a high priority. Future studies may better promote the benefits of participation (free CBCs and diagnostics) or provide an incentive for the veterinarian. Interestingly, follow-up with clinic-recruited control owners for the prospective study had a low success rate. Because veterinarians recruited this population, there was less contact between owners and researchers. We suspect that not having an established relationship reduced their interest in participating in follow-up studies. Additionally, these owners may have been less invested initially because they were randomly recruited as opposed to a vested interest driving them to volunteer (as is the case in CLHP-recruited owners).

CLHP-recruitment was largely successful, albeit time-intensive. The initial participation invitation sent to these owners resulted in a surge of calls and emails and subsequent surge of samples that overloaded our laboratory. Ideally, we would spread this out in the future through staging the email blasts over a multi-week period. Recruitment among this population required pre-screening to ensure dogs were the correct age and lymphoma-free (other than TZL), followed by contacting the owners to relay enrollment information (provide a survey link and sample shipping information). Additional contact with owners was often required to coordinate sample shipment and send reminders for both the survey and blood sample. This group of owners were generally quite invested in participating in this research project, which required a

greater researcher time commitment, but had many benefits. Specifically, many of these owners recruited other participants, further increasing enrollment. Additionally, there has been a high willingness to participate in follow-up studies. As we do not have access to the original database, it is unknown what percent of owners who received the initial callout ultimately contacted us. However, of those who expressed interest in participating, 70% completed enrollment.

There are strengths and limitations to each recruitment strategy. Clinic-recruited controls are likely more representative of the case source population and thus are superior from an epidemiological perspective. However, recruitment from this source was less successful and ultimately did not provide enough samples for adequately analysis. This recruitment strategy placed a large burden on veterinarians who are busy and have less incentive to help. If we could reduce veterinarian burden and increase enrollment rate, this would likely be the preferred recruitment strategy. CLHP-recruited controls represent a great source of a large number of geographically distributed dogs. These owners were very responsive and invested in the research topic. However, this population may not be representative of the case source population, thereby causing a selection bias. This bias could be mitigated to some degree if we could create a variable that quantified systematic differences due to recruitment strategy. As use of the CLHP database increases, evaluation of demographics and behaviors of these owners may shed light on a way to quantify these differences.

## Overall conclusions

*Summary of conclusions in the context of lymphoma research*

Overall, our research contributes to the body of evidence that genetic and antigenic risk factors play a role in both B-CLL and TZL pathogenesis. We identified a strong breed predisposition for B-CLL and specific genetic risk factors for TZL. Additionally, we identified multiple inflammatory factors associated with TZL, providing evidence for a role of chronic antigen stimulation. Although not part of this dissertation, continued work in our laboratory has identified restricted $IgV_H$ gene usage in canine B-CLL, which also suggests a role of chronic antigen stimulation.

Findings from this dissertation can be used to inform future mechanistic studies, which can subsequently benefit all aspects of lymphoma research, including disease prevention, screening and early diagnosis, identification of prognostic factors, and development of novel treatments. In particular, the identification of a precursor T cell neoplasm is novel, and future evaluation of subclinical dogs may provide valuable insights into T cell lymphoma pathogenesis. This would not only help the large numbers of dogs who are affected by T cell lymphoma, but also the human populations in which epidemiologic studies have been challenging to conduct due to low numbers of cases.

*Summary of conclusions in the context of the canine cancer model*

This dissertation sheds light on novel uses of the canine cancer model and the utility of epidemiologic studies in dogs. Our research demonstrated multiple ways to relatively quickly and inexpensively recruit large canine populations. We showed a high response rate in our

case-control study, highlighting the importance of such research to owners and their investment in participating. Despite the fact that there is less infrastructure to conduct large studies than we have in human populations, they are feasible to implement. The value of canine research extends beyond furthering veterinary medicine, and is particularly useful for human diseases that are comparatively rare and thus more difficult to study.

# REFERENCES


1. Fais F, Ghiotto F, Hashimoto S, Sellars B, Valetto A, Allen SL, et al. Chronic lymphocytic leukemia B cells express restricted sets of mutated and unmutated antigen receptors. J Clin Invest 1998; 102:1515-25.
2. Tobin G, Thunberg U, Johnson A, Eriksson I, Söderberg O, Karlsson K, et al. Chronic lymphocytic leukemias utilizing the VH3-21 gene display highly restricted Vlambda2-14 gene use and homologous CDR3s: implicating recognition of a common antigen epitope. Blood 2003; 101:4952-7.
3. Kipps TJ, Tomhave E, Pratt LF, Duffy S, Chen PP, Carson DA. Developmentally restricted immunoglobulin heavy chain variable region gene expressed at high frequency in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A 1989; 86:5913-7.
4. Tobin G, Thunberg U, Karlsson K, Murray F, Laurell A, Willander K, et al. Subsets with restricted immunoglobulin gene rearrangement features indicate a role for antigen selection in the development of chronic lymphocytic leukemia. Blood 2004; 104:2879-85.
5. Bachy E, Urb M, Chandra S, Robinot R, Bricard G, de Bernard S, et al. CD1d-restricted peripheral T cell lymphoma in mice and humans. J Exp Med 2016; 213:841-57.
6. Seelig DM, Avery P, Webb T, Yoshimoto J, Bromberek J, Ehrhart EJ, et al. Canine T-zone lymphoma: unique immunophenotypic features, outcome, and population characteristics. J Vet Intern Med 2014; 28:878-86.
7. Valli VE, San Myint M, Barthel A, Bienzle D, Caswell J, Colbatzky F, et al. Classification of canine malignant lymphomas according to the World Health Organization criteria. Vet Pathol 2011; 48:198-211.
8. Valli VE, Vernau W, de Lorimier LP, Graham PS, Moore PF. Canine indolent nodular lymphoma. Vet Pathol 2006; 43:241-56.
9. American Cancer Society. Cancer Facts & Figures 2017. Atlanta: American Cancer Society, 2017.
10. Goldin LR, Björkholm M, Kristinsson SY, Turesson I, Landgren O. Elevated risk of chronic lymphocytic leukemia and other indolent non-Hodgkin's lymphomas among relatives of patients with chronic lymphocytic leukemia. Haematologica 2009; 94:647-53.
11. CanineHD BeadChip. In: Illumina I, ed. Pub. No. 370-2009-018. San Diego, CA, 2009.
12. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010; 7:248-9.
13. Olzmann JA, Kopito RR, Christianson JC. The mammalian endoplasmic reticulum-associated degradation system. Cold Spring Harb Perspect Biol 2013; 5.
14. Diril MK, Wienisch M, Jung N, Klingauf J, Haucke V. Stonin 2 is an AP-2-dependent endocytic sorting adaptor for synaptotagmin internalization and recycling. Dev Cell 2006; 10:233-44.
15. Society AC. Cancer Facts & Figures 2016. Atlanta: American Cancer Society, 2016.

16. Dorn CR, Taylor DO, Schneider R, Hibbard HH, Klauber MR. Survey of animal neoplasms in Alameda and Contra Costa Counties, California. II. Cancer morbidity in dogs and cats from Alameda County. J Natl Cancer Inst 1968; 40:307-18.
17. MacVean DW, Monlux AW, Anderson PS, Silberg SL, Roszel JF. Frequency of canine and feline tumors in a defined population. Vet Pathol 1978; 15:700-15.
18. Dobson JM, Samuel S, Milstein H, Rogers K, Wood JL. Canine neoplasia in the UK: estimates of incidence rates from a population of insured dogs. J Small Anim Pract 2002; 43:240-6.
19. Howlader N, Noone A, Krapcho M, Miller D, Bishop K, Altekruse S, et al. SEER Cancer Statistics Review, 1975-2013. National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2013/, based on November 2015 SEER data submission, posted to the SEER web site, April 2016.
20. Swerdlow S, Campo E, Harris N, al. e, eds. WHO Classification of Tumors of Hematopoietic and Lymphoid Tissues. Lyon, France: International Agency for Research on Cancer; 2008.
21. Perry AM, Diebold J, Nathwani BN, MacLennan KA, Müller-Hermelink HK, Bast M, et al. Non-Hodgkin lymphoma in the developing world: review of 4539 cases from the International Non-Hodgkin Lymphoma Classification Project. Haematologica 2016.
22. Armitage JO, Weisenburger DD. New approach to classifying non-Hodgkin's lymphomas: clinical features of the major histologic subtypes. Non-Hodgkin's Lymphoma Classification Project. J Clin Oncol 1998; 16:2780-95.
23. Cabanillas F, Velasquez WS, Hagemeister FB, McLaughlin P, Redman JR. Clinical, biologic, and histologic features of late relapses in diffuse large cell lymphoma. Blood 1992; 79:1024-8.
24. Bastion Y, Sebban C, Berger F, Felman P, Salles G, Dumontet C, et al. Incidence, predictive factors, and outcome of lymphoma transformation in follicular lymphoma patients. J Clin Oncol 1997; 15:1587-94.
25. National Comprehensive Cancer Network. Non Hodgkin's Lymphoma (Version 3.2016). Available from https://www.nccn.org/professionals/physician_gls/pdf/nhl.pdf.
26. Keller ET, MacEwen EG, Rosenthal RC, Helfand SC, Fox LE. Evaluation of prognostic factors and sequential combination chemotherapy with doxorubicin for canine lymphoma. J Vet Intern Med 1993; 7:289-95.
27. Garrett LD, Thamm DH, Chun R, Dudley R, Vail DM. Evaluation of a 6-month chemotherapy protocol with no maintenance therapy for dogs with lymphoma. J Vet Intern Med 2002; 16:704-9.
28. Flood-Knapik KE, Durham AC, Gregor TP, Sánchez MD, Durney ME, Sorenmo KU. Clinical, histopathological and immunohistochemical characterization of canine indolent lymphoma. Vet Comp Oncol 2012.
29. Khanna C, Lindblad-Toh K, Vail D, London C, Bergman P, Barber L, et al. The dog as a cancer model. Nat Biotechnol 2006; 24:1065-6.
30. Fürdös I, Fazekas J, Singer J, Jensen-Jarolim E. Translating clinical trials from human to veterinary oncology and back. J Transl Med 2015; 13:265.

31. Roth HS, Hergenrother PJ. Derivatives of Procaspase-Activating Compound 1 (PAC-1) and their Anticancer Activities. Curr Med Chem 2016; 23:201-41.
32. Guy MK, Page RL, Jensen WA, Olson PN, Haworth JD, Searfoss EE, et al. The Golden Retriever Lifetime Study: establishing an observational cohort study with translational relevance for human health. Philos Trans R Soc Lond B Biol Sci 2015; 370.
33. Kirkness EF, Bafna V, Halpern AL, Levy S, Remington K, Rusch DB, et al. The dog genome: survey sequencing and comparative analysis. Science 2003; 301:1898-903.
34. Ostrander EA, Galibert F, Patterson DF. Canine genetics comes of age. Trends Genet 2000; 16:117-24.
35. Patterson DF. Companion animal medicine in the age of medical genetics. J Vet Intern Med 2000; 14:1-9.
36. Chase K, Carrier DR, Adler FR, Jarvik T, Ostrander EA, Lorentzen TD, et al. Genetic basis for systems of skeletal quantitative traits: principal component analysis of the canid skeleton. Proc Natl Acad Sci U S A 2002; 99:9930-5.
37. Ostrander EA. Franklin H. Epstein Lecture. Both ends of the leash--the human links to good dogs with bad genes. N Engl J Med 2012; 367:636-46.
38. Lindblad-Toh K, Wade CM, Mikkelsen TS, Karlsson EK, Jaffe DB, Kamal M, et al. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 2005; 438:803-19.
39. Sutter NB, Eberle MA, Parker HG, Pullar BJ, Kirkness EF, Kruglyak L, et al. Extensive and breed-specific linkage disequilibrium in Canis familiaris. Genome Res 2004; 14:2388-96.
40. Ito D, Frantz AM, Modiano JF. Canine lymphoma as a comparative model for human non-Hodgkin lymphoma: recent progress and applications. Vet Immunol Immunopathol 2014; 159:192-201.
41. Davis BW, Ostrander EA. Domestic dogs and cancer research: a breed-based genomics approach. ILAR J 2014; 55:59-68.
42. Karlsson EK, Baranowska I, Wade CM, Salmon Hillbertz NH, Zody MC, Anderson N, et al. Efficient mapping of mendelian traits in dogs through genome-wide association. Nat Genet 2007; 39:1321-8.
43. Shearin AL, Ostrander EA. Leading the way: canine models of genomics and disease. Dis Model Mech 2010; 3:27-34.
44. Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT. Data quality control in genetic case-control association studies. Nat Protoc 2010; 5:1564-73.
45. PLINK [1.9]. Available from https://www.cog-genomics.org/plink2.
46. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011; 88:76-82.
47. Brønden LB, Flagstad A, Kristensen AT. Veterinary cancer registries in companion animal cancer: a review. Vet Comp Oncol 2007; 5:133-44.
48. Hayes HM, Tarone RE, Cantor KP, Jessen CR, McCurnin DM, Richardson RC. Case-control study of canine malignant lymphoma: positive association with dog owner's use of 2,4-dichlorophenoxyacetic acid herbicides. J Natl Cancer Inst 1991; 83:1226-31.

49. Keller ET. Immune-mediated disease as a risk factor for canine lymphoma. Cancer 1992; 70:2334-7.
50. Villamil JA, Henry CJ, Hahn AW, Bryan JN, Tyler JW, Caldwell CW. Hormonal and sex impact on the epidemiology of canine lymphoma. J Cancer Epidemiol 2009; 2009:591753.
51. Reif JS, Lower KS, Ogilvie GK. Residential exposure to magnetic fields and risk of canine lymphoma. Am J Epidemiol 1995; 141:352-9.
52. Takashima-Uebelhoer BB, Barber LG, Zagarins SE, Procter-Gray E, Gollenberg AL, Moore AS, et al. Household chemical exposures and the risk of canine malignant lymphoma, a model for human non-Hodgkin's lymphoma. Environ Res 2012; 112:171-6.
53. Marconato L, Leo C, Girelli R, Salvi S, Abramo F, Bettini G, et al. Association between waste management and cancer in companion animals. J Vet Intern Med 2009; 23:564-9.
54. Duncan AW, Marr HS, Birkenheuer AJ, Maggi RG, Williams LE, Correa MT, et al. Bartonella DNA in the blood and lymph nodes of Golden Retrievers with lymphoma and in healthy controls. J Vet Intern Med 2008; 22:89-95.
55. Edwards DS, Henley WE, Harding EF, Dobson JM, Wood JL. Breed incidence of lymphoma in a UK population of insured dogs. Vet Comp Oncol 2003; 1:200-6.
56. Merlo DF, Rossi L, Pellegrino C, Ceppi M, Cardellino U, Capurro C, et al. Cancer incidence in pet dogs: findings of the Animal Tumor Registry of Genoa, Italy. J Vet Intern Med 2008; 22:976-84.
57. BERKSON J. Limitations of the application of fourfold table analysis to hospital data. Biometrics 1946; 2:47-53.
58. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology 2003; 14:300-6.
59. Carlo GL, Cole P, Miller AB, Munro IC, Solomon KR, Squire RA. Review of a study reporting an association between 2,4-dichlorophenoxyacetic acid and canine malignant lymphoma: report of an expert panel. Regul Toxicol Pharmacol 1992; 16:245-52.
60. Hayes HM, Tarone RE, Cantor KP. On the association between canine malignant lymphoma and opportunity for exposure to 2,4-dichlorophenoxyacetic acid. Environ Res 1995; 70:119-25.
61. Torres de la Riva G, Hart BL, Farver TB, Oberbauer AM, Messam LL, Willits N, et al. Neutering dogs: effects on joint disorders and cancers in golden retrievers. PLoS One 2013; 8:e55937.
62. Hartge P, Smith MT. Environmental and behavioral factors and the risk of non-Hodgkin lymphoma. Cancer Epidemiol Biomarkers Prev 2007; 16:367-8.
63. Morton LM, Wang SS, Cozen W, Linet MS, Chatterjee N, Davis S, et al. Etiologic heterogeneity among non-Hodgkin lymphoma subtypes. Blood 2008; 112:5150-60.
64. Anderson LA, Gadalla S, Morton LM, Landgren O, Pfeiffer R, Warren JL, et al. Population-based study of autoimmune conditions and the risk of specific lymphoid malignancies. Int J Cancer 2009; 125:398-405.
65. Baecklund E, Iliadou A, Askling J, Ekbom A, Backlin C, Granath F, et al. Association of chronic inflammation, not its treatment, with increased lymphoma risk in rheumatoid arthritis. Arthritis Rheum 2006; 54:692-701.

66. Clarke CA, Morton LM, Lynch C, Pfeiffer RM, Hall EC, Gibson TM, et al. Risk of lymphoma subtypes after solid organ transplantation in the United States. Br J Cancer 2013; 109:280-8.
67. Quinlan SC, Morton LM, Pfeiffer RM, Anderson LA, Landgren O, Warren JL, et al. Increased risk for lymphoid and myeloid neoplasms in elderly solid-organ transplant recipients. Cancer Epidemiol Biomarkers Prev 2010; 19:1229-37.
68. Grulich AE, van Leeuwen MT, Falster MO, Vajdic CM. Incidence of cancers in people with HIV/AIDS compared with immunosuppressed transplant recipients: a meta-analysis. Lancet 2007; 370:59-67.
69. Dal Maso L, Franceschi S. Epidemiology of non-Hodgkin lymphomas and other haemolymphopoietic neoplasms in people with AIDS. Lancet Oncol 2003; 4:110-9.
70. Coté TR, Biggar RJ, Rosenberg PS, Devesa SS, Percy C, Yellin FJ, et al. Non-Hodgkin's lymphoma among people with AIDS: incidence, presentation and public health burden. AIDS/Cancer Study Group. Int J Cancer 1997; 73:645-50.
71. Biggar RJ, Engels EA, Frisch M, Goedert JJ, Group ACMRS. Risk of T-cell lymphomas in persons with AIDS. J Acquir Immune Defic Syndr 2001; 26:371-6.
72. Slager SL, Benavente Y, Blair A, Vermeulen R, Cerhan JR, Costantini AS, et al. Medical history, lifestyle, family history, and occupational risk factors for chronic lymphocytic leukemia/small lymphocytic lymphoma: the InterLymph Non-Hodgkin Lymphoma Subtypes Project. J Natl Cancer Inst Monogr 2014; 2014:41-51.
73. Teske E, de Vos JP, Egberink HF, Vos JH. Clustering in canine malignant lymphoma. Vet Q 1994; 16:134-6.
74. Modiano JF, Breen M, Burnett RC, Parker HG, Inusah S, Thomas R, et al. Distinct B-cell and T-cell lymphoproliferative disease prevalence among dog breeds indicates heritable risk. Cancer Res 2005; 65:5654-61.
75. Tonomura N, Elvers I, Thomas R, Megquier K, Turner-Maier J, Howald C, et al. Genome-wide association study identifies shared risk loci common to two malignancies in golden retrievers. PLoS Genet 2015; 11:e1004922.
76. Elvers I, Turner-Maier J, Swofford R, Koltookian M, Johnson J, Stewart C, et al. Exome sequencing of lymphomas from three dog breeds reveals somatic mutation patterns reflecting genetic background. Genome Res 2015; 25:1634-45.
77. Society AC. Cancer Facts & Figures 2014. Atlanta: American Cancer Society, 2014.
78. Dighiero G, Hamblin TJ. Chronic lymphocytic leukaemia. Lancet 2008; 371:1017-29.
79. Hallek M, Cheson BD, Catovsky D, Caligaris-Cappio F, Dighiero G, Döhner H, et al. Guidelines for the diagnosis and treatment of chronic lymphocytic leukemia: a report from the International Workshop on Chronic Lymphocytic Leukemia updating the National Cancer Institute-Working Group 1996 guidelines. Blood 2008; 111:5446-56.
80. Weiss NS. Geographical Variation in the Incidence of the Leukemias and Lymphomas. National Cancer Institute monograph 1979:139 - 42.
81. Dores GM, Anderson WF, Curtis RE, Landgren O, Ostroumova E, Bluhm EC, et al. Chronic lymphocytic leukaemia and small lymphocytic lymphoma: overview of the descriptive epidemiology. Br J Haematol 2007; 139:809-19.

82. Crowther-Swanepoel D, Broderick P, Di Bernardo MC, Dobbins SE, Torres M, Mansouri M, et al. Common variants at 2q37.3, 8q24.21, 15q21.3 and 16q24.1 influence chronic lymphocytic leukemia risk. Nat Genet 2010; 42:132-6.
83. Slager SL, Skibola CF, Di Bernardo MC, Conde L, Broderick P, McDonnell SK, et al. Common variation at 6p21.31 (BAK1) influences the risk of chronic lymphocytic leukemia. Blood 2012; 120:843-6.
84. Berndt SI, Skibola CF, Joseph V, Camp NJ, Nieters A, Wang Z, et al. Genome-wide association study identifies multiple risk loci for chronic lymphocytic leukemia. Nat Genet 2013; 45:868-76.
85. Di Bernardo MC, Crowther-Swanepoel D, Broderick P, Webb E, Sellick G, Wild R, et al. A genome-wide association study identifies six susceptibility loci for chronic lymphocytic leukemia. Nat Genet 2008; 40:1204-10.
86. Speedy HE, Di Bernardo MC, Sava GP, Dyer MJ, Holroyd A, Wang Y, et al. A genome-wide association study identifies multiple susceptibility loci for chronic lymphocytic leukemia. Nat Genet 2014; 46:56-60.
87. Vanura K, Le T, Esterbauer H, Späth F, Porpaczy E, Shehata M, et al. Autoimmune conditions and chronic infections in chronic lymphocytic leukemia patients at diagnosis are associated with unmutated IgVH genes. Haematologica 2008; 93:1912-6.
88. Leifer CE, Matus RE. Chronic lymphocytic leukemia in the dog: 22 cases (1974-1984). J Am Vet Med Assoc 1986; 189:214-7.
89. Vernau W, Moore PF. An immunophenotypic study of canine leukemias and preliminary assessment of clonality by polymerase chain reaction. Vet Immunol Immunopathol 1999; 69:145-64.
90. Adam F, Villiers E, Watson S, Coyne K, Blackwood L. Clinical pathological and epidemiological assessment of morphologically and immunologically confirmed canine leukaemia. Vet Comp Oncol 2009; 7:181-95.
91. Comazzi S, Gelain ME, Martini V, Riondato F, Miniscalco B, Marconato L, et al. Immunophenotype predicts survival time in dogs with chronic lymphocytic leukemia. Journal of Veterinary Internal Medicine 2011; 25:100-6.
92. Adam F, Villiers E, Watson S, Coyne K, Blackwood L. Clinical pathological and epidemiological assessment of morphologically and immunologically confirmed canine leukaemia. Veterinary & Comparative Oncology 2009; 7:181-95.
93. Williams MJ, Avery AC, Lana SE, Hillers KR, Bachand AM, Avery PR. Canine lymphoproliferative disease characterized by lymphocytosis: immunophenotypic markers of prognosis. J Vet Intern Med 2008; 22:596-601.
94. Linet MS, Van Natta ML, Brookmeyer R, Khoury MJ, McCaffrey LD, Humphrey RL, et al. Familial cancer history and chronic lymphocytic leukemia. A case-control study. Am J Epidemiol 1989; 130:655-64.
95. Sellick GS, Catovsky D, Houlston RS. Familial chronic lymphocytic leukemia. Semin Oncol 2006; 33:195-201.
96. Anderson JR, Armitage JO, Weisenburger DD. Epidemiology of the non-Hodgkin's lymphomas: distributions of the major subtypes differ by geographic locations. Non-Hodgkin's Lymphoma Classification Project. Ann Oncol 1998; 9:717-20.

97. Vose J, Armitage J, Weisenburger D, International T-Cell Lymphoma Project. International peripheral T-cell and natural killer/T-cell lymphoma study: pathology findings and clinical outcomes. J Clin Oncol 2008; 26:4124-30.
98. Weisenburger DD, Savage KJ, Harris NL, Gascoyne RD, Jaffe ES, MacLennan KA, et al. Peripheral T-cell lymphoma, not otherwise specified: a report of 340 cases from the International Peripheral T-cell Lymphoma Project. Blood 2011; 117:3402-8.
99. Sonnen R, Schmidt WP, Müller-Hermelink HK, Schmitz N. The International Prognostic Index determines the outcome of patients with nodal mature T-cell lymphomas. Br J Haematol 2005; 129:366-72.
100. Savage KJ, Ferreri AJ, Zinzani PL, Pileri SA. Peripheral T-cell lymphoma--not otherwise specified. Crit Rev Oncol Hematol 2011; 79:321-9.
101. Foss FM, Zinzani PL, Vose JM, Gascoyne RD, Rosen ST, Tobinai K. Peripheral T-cell lymphoma. Blood 2011; 117:6756-67.
102. Shapira MY, Caspi O, Amir G, Zlotogorski A, Naparstek Y. Gastric-Mucocutaneous gammadelta T cell lymphoma: possible association with Epstein-Barr virus? Leuk Lymphoma 1999; 35:397-401.
103. Wang SS, Flowers CR, Kadin ME, Chang ET, Hughes AM, Ansell SM, et al. Medical history, lifestyle, family history, and occupational risk factors for peripheral T-cell lymphomas: the InterLymph Non-Hodgkin Lymphoma Subtypes Project. J Natl Cancer Inst Monogr 2014; 2014:66-75.
104. Piccaluga PP, Agostinelli C, Tripodo C, Gazzola A, Bacci F, Sabattini E, et al. Peripheral T-cell lymphoma classification: the matter of cellular derivation. Expert Rev Hematol 2011; 4:415-25.
105. de Leval L, Rickman DS, Thielen C, Reynies A, Huang YL, Delsol G, et al. The gene expression profile of nodal peripheral T-cell lymphoma demonstrates a molecular link between angioimmunoblastic T-cell lymphoma (AITL) and follicular helper T (TFH) cells. Blood 2007; 109:4952-63.
106. Romagnani S. Biology of human TH1 and TH2 cells. J Clin Immunol 1995; 15:121-9.
107. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. CA Cancer J Clin 2014; 64:9-29.
108. Comazzi S, Gelain ME, Martini V, Riondato F, Miniscalco B, Marconato L, et al. Immunophenotype predicts survival time in dogs with chronic lymphocytic leukemia. J Vet Intern Med 2011; 25:100-6.
109. Rai KR, Sawitsky A, Cronkite EP, Chanana AD, Levy RN, Pasternack BS. Clinical staging of chronic lymphocytic leukemia. Blood 1975; 46:219-34.
110. Hamblin TJ, Davis Z, Gardiner A, Oscier DG, Stevenson FK. Unmutated Ig V(H) genes are associated with a more aggressive form of chronic lymphocytic leukemia. Blood 1999; 94:1848-54.
111. Huang PY, Best OG, Almazi JG, Belov L, Davis ZA, Majid A, et al. Cell surface phenotype profiles distinguish stable and progressive chronic lymphocytic leukemia. Leuk Lymphoma 2014; 55:2085-92.
112. Rao S, Lana S, Eickhoff J, Marcus E, Avery PR, Morley PS, et al. Class II major histocompatibility complex expression and cell size independently predict survival in canine B-cell lymphoma. J Vet Intern Med 2011; 25:1097-105.

113. Tretter T, Venigalla RK, Eckstein V, Saffrich R, Sertel S, Ho AD, et al. Induction of CD4+ T-cell anergy and apoptosis by activated human B cells. Blood 2008; 112:4555-64.
114. Amu S, Tarkowski A, Dörner T, Bokarewa M, Brisslert M. The human immunomodulatory CD25+ B cell population belongs to the memory B cell pool. Scand J Immunol 2007; 66:77-86.
115. Brisslert M, Bokarewa M, Larsson P, Wing K, Collins LV, Tarkowski A. Phenotypic and functional characterization of human CD25+ B cells. Immunology 2006; 117:548-57.
116. Serra M, Freeman KP, Campora C, Sacchini F. Establishment of canine hematology reference intervals for the Sysmex XT-2000iV hematology analyzer using a blood donor database. Vet Clin Pathol 2012; 41:207-15.
117. Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, et al. A single IGF1 allele is a major determinant of small size in dogs. Science 2007; 316:112-5.
118. Rimbault M, Beale HC, Schoenebeck JJ, Hoopes BC, Allen JJ, Kilroy-Glynn P, et al. Derived variants at six genes explain nearly half of size reduction in dog breeds. Genome Res 2013; 23:1985-95.
119. Yee KW, O'Brien SM. Chronic lymphocytic leukemia: diagnosis and treatment. Mayo Clin Proc 2006; 81:1105-29.
120. Tasca S, Carli E, Caldin M, Menegazzo L, Furlanello T, Gallego LS. Hematologic abnormalities and flow cytometric immunophenotyping results in dogs with hematopoietic neoplasia: 210 cases (2002-2006). Vet Clin Pathol 2009; 38:2-12.
121. Vezzali E, Parodi AL, Marcato PS, Bettini G. Histopathologic classification of 171 cases of canine and feline non-Hodgkin lymphoma according to the WHO. Vet Comp Oncol 2010; 8:38-49.
122. Pastor M, Chalvet-Monfray K, Marchal T, Keck G, Magnol JP, Fournel-Fleury C, et al. Genetic and environmental risk indicators in canine non-Hodgkin's lymphomas: breed associations and geographic distribution of 608 cases diagnosed throughout France over 1 year. J Vet Intern Med 2009; 23:301-10.
123. Rüdiger T, Weisenburger DD, Anderson JR, Armitage JO, Diebold J, MacLennan KA, et al. Peripheral T-cell lymphoma (excluding anaplastic large-cell lymphoma): results from the Non-Hodgkin's Lymphoma Classification Project. Ann Oncol 2002; 13:140-9.
124. Iqbal J, Wright G, Wang C, Rosenwald A, Gascoyne RD, Weisenburger DD, et al. Gene expression signatures delineate biological and prognostic subgroups in peripheral T-cell lymphoma. Blood 2014; 123:2915-23.
125. Valli V. Mature (peripheral) nodal T-cell (T-zone) lymphoma. In: Veterinary Comparative Hematopathology. Ames, IA: Blackwell; 2007: 294-302.
126. Vaysse A, Ratnakumar A, Derrien T, Axelsson E, Rosengren Pielberg G, Sigurdsson S, et al. Identification of genomic regions associated with phenotypic variation between dog breeds using selection mapping. PLoS Genet 2011; 7:e1002316.
127. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015; 4:7.
128. Arendt ML, Melin M, Tonomura N, Koltookian M, Courtay-Cahen C, Flindall N, et al. Genome-Wide Association Study of Golden Retrievers Identifies Germ-Line Risk Factors Predisposing to Mast Cell Tumours. PLoS Genet 2015; 11:e1005647.

129. Thomson R, McWhirter R. Adjusting for Familial Relatedness in the Analysis of GWAS Data. Methods Mol Biol 2017; 1526:175-90.
130. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 2005; 21:263-5.
131. FastQC: a quality control tool for high throughput sequence data. Babraham Bioinformatics. Available from http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/.
132. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009; 25:2078-9.
133. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011; 27:2987-93.
134. Picard. Available from http://picard.sourceforge.net/.
135. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20:1297-303.
136. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. q-bio.GN 2013:arXiv :1303.3997v1.
137. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol 2011; 29:24-6.
138. Cingolani P, Platts A, Wang lL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin) 2012; 6:80-92.
139. Toole BP. Hyaluronan in morphogenesis. Semin Cell Dev Biol 2001; 12:79-87.
140. Tian X, Azpurua J, Hine C, Vaidya A, Myakishev-Rempel M, Ablaeva J, et al. High-molecular-mass hyaluronan mediates the cancer resistance of the naked mole rat. Nature 2013; 499:346-9.
141. Girish KS, Kemparaju K. The magic glue hyaluronan and its eraser hyaluronidase: a biological overview. Life Sci 2007; 80:1921-43.
142. Csóka AB, Scherer SW, Stern R. Expression analysis of six paralogous human hyaluronidase genes clustered on chromosomes 3p21 and 7q31. Genomics 1999; 60:356-61.
143. Cherr GN, Yudin AI, Overstreet JW. The dual functions of GPI-anchored PH-20: hyaluronidase and intracellular signaling. Matrix Biol 2001; 20:515-25.
144. Zhang H, Martin-DeLeon PA. Mouse Spam1 (PH-20) is a multifunctional protein: evidence for its expression in the female reproductive tract. Biol Reprod 2003; 69:446-54.
145. Oettl M, Hoechstetter J, Asen I, Bernhardt G, Buschauer A. Comparative characterization of bovine testicular hyaluronidase and a hyaluronate lyase from Streptococcus agalactiae in pharmaceutical preparations. Eur J Pharm Sci 2003; 18:267-77.
146. Beech DJ, Madan AK, Deng N. Expression of PH-20 in normal and neoplastic breast tissue. J Surg Res 2002; 103:203-7.

147. Ozcan U, Cao Q, Yilmaz E, Lee AH, Iwakoshi NN, Ozdelen E, et al. Endoplasmic reticulum stress links obesity, insulin action, and type 2 diabetes. Science 2004; 306:457-61.
148. Gao B, Lee SM, Chen A, Zhang J, Zhang DD, Kannan K, et al. Synoviolin promotes IRE1 ubiquitination and degradation in synovial fibroblasts from mice with collagen-induced arthritis. EMBO Rep 2008; 9:480-5.
149. Imai Y, Soda M, Inoue H, Hattori N, Mizuno Y, Takahashi R. An unfolded putative transmembrane polypeptide, which can lead to endoplasmic reticulum stress, is a substrate of Parkin. Cell 2001; 105:891-902.
150. Kaneko M, Koike H, Saito R, Kitamura Y, Okuma Y, Nomura Y. Loss of HRD1-mediated protein degradation causes amyloid precursor protein accumulation and amyloid-beta generation. J Neurosci 2010; 30:3924-32.
151. Artavanis-Tsakonas S, Rand MD, Lake RJ. Notch signaling: cell fate control and signal integration in development. Science 1999; 284:770-6.
152. Cattaneo M, Orlandini S, Beghelli S, Moore PS, Sorio C, Bonora A, et al. SEL1L expression in pancreatic adenocarcinoma parallels SMAD4 expression and delays tumor growth in vitro and in vivo. Oncogene 2003; 22:6359-68.
153. Orlandi R, Cattaneo M, Troglio F, Casalini P, Ronchini C, Ménard S, et al. SEL1L expression decreases breast tumor cell aggressiveness in vivo and in vitro. Cancer Res 2002; 62:567-74.
154. Cattaneo M, Baronchelli S, Schiffer D, Mellai M, Caldera V, Saccani GJ, et al. Down-modulation of SEL1L, an unfolded protein response and endoplasmic reticulum-associated degradation protein, sensitizes glioma stem cells to the cytotoxic effect of valproic acid. J Biol Chem 2014; 289:2826-38.
155. Roche PA, Furuta K. The ins and outs of MHC class II-mediated antigen processing and presentation. Nat Rev Immunol 2015; 15:203-16.
156. Platt CD, Ma JK, Chalouni C, Ebersold M, Bou-Reslan H, Carano RA, et al. Mature dendritic cells use endocytic receptors to capture and present antigens. Proc Natl Acad Sci U S A 2010; 107:4287-92.
157. Imai J, Hasegawa H, Maruya M, Koyasu S, Yahara I. Exogenous antigens are processed through the endoplasmic reticulum-associated degradation (ERAD) in cross-presentation by dendritic cells. Int Immunol 2005; 17:45-53.
158. Ackerman AL, Giodini A, Cresswell P. A role for the endoplasmic reticulum protein retrotranslocation machinery during crosspresentation by dendritic cells. Immunity 2006; 25:607-17.
159. Giodini A, Rahner C, Cresswell P. Receptor-mediated phagocytosis elicits cross-presentation in nonprofessional antigen-presenting cells. Proc Natl Acad Sci U S A 2009; 106:3324-9.
160. Park HY, Lee SB, Yoo HY, Kim SJ, Kim WS, Kim JI, et al. Whole-exome and transcriptome sequencing of refractory diffuse large B-cell lymphoma. Oncotarget 2016; 7:86433-45.
161. Graham PA, Refsal KR, Nachreiner RF. Etiopathologic findings of canine hypothyroidism. Vet Clin North Am Small Anim Pract 2007; 37:617-31, v.

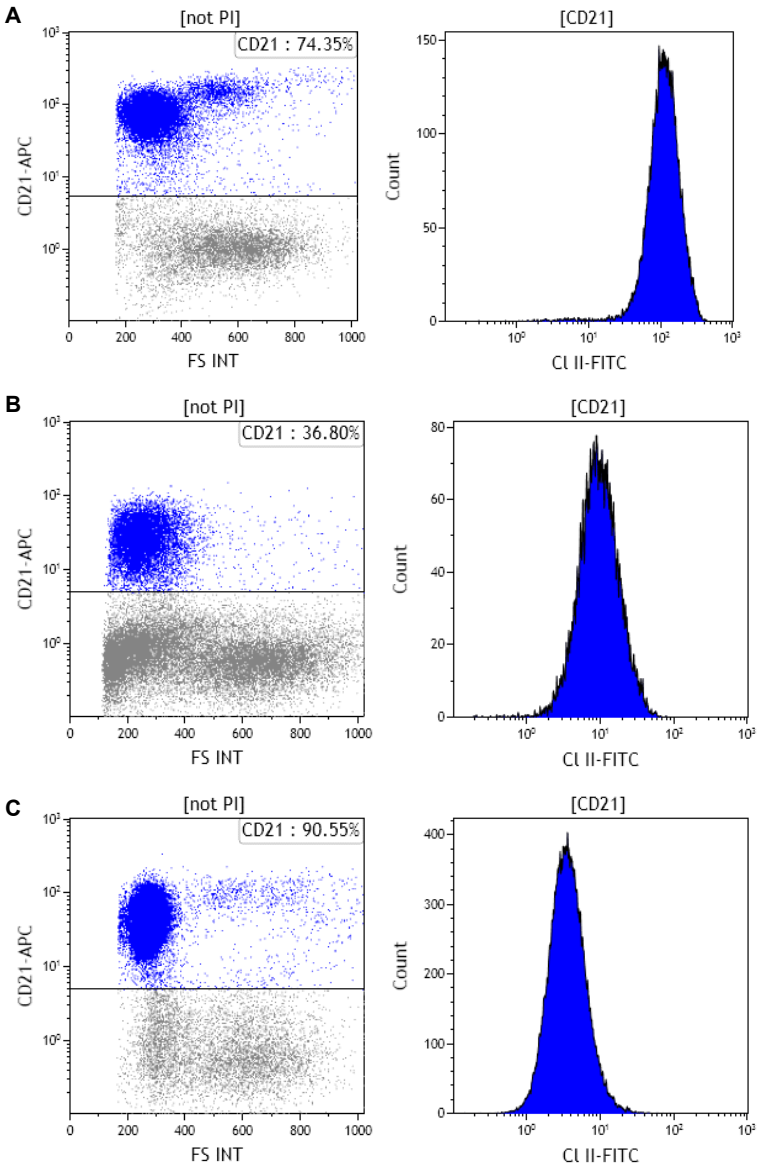
162. Dietrich JW, Landgrafe G, Fotiadou EH. TSH and Thyrotropic Agonists: Key Actors in Thyroid Homeostasis. J Thyroid Res 2012; 2012:351864.
163. Gereben B, Zavacki AM, Ribich S, Kim BW, Huang SA, Simonides WS, et al. Cellular and molecular basis of deiodinase-regulated thyroid hormone signaling. Endocr Rev 2008; 29:898-938.
164. Arrojo E Drigo R, Fonseca TL, Castillo M, Salathe M, Simovic G, Mohácsik P, et al. Endoplasmic reticulum stress decreases intracellular thyroid hormone activation via an eIF2a-mediated decrease in type 2 deiodinase synthesis. Mol Endocrinol 2011; 25:2065-75.
165. Hernandez JD, Baum LG. Ah, sweet mystery of death! Galectins and control of cell fate. Glycobiology 2002; 12:127R-36R.
166. Holly VL, Sandmeyer LS, Bauer BS, Verges L, Grahn BH. Golden retriever cystic uveal disease: a longitudinal study of iridociliary cysts, pigmentary uveitis, and pigmentary/cystic glaucoma over a decade in western Canada. Vet Ophthalmol 2016; 19:237-44.
167. Ocular Disorders Presumed to be Inherited in Purebred Dogs. 2nd ed. Phoenix, AZ: Genetic Committee of the American College of Veterinary Ophthalmologists; 1996.
168. Deehr AJ, Dubielzig RR. A histopathological study of iridociliary cysts and glaucoma in Golden Retrievers. Vet Ophthalmol 1998; 1:153-8.
169. Maroon JC, Bost JW. Omega-3 fatty acids (fish oil) as an anti-inflammatory: an alternative to nonsteroidal anti-inflammatory drugs for discogenic pain. Surg Neurol 2006; 65:326-31.
170. Smith WL. Cyclooxygenases, peroxide tone and the allure of fish oil. Curr Opin Cell Biol 2005; 17:174-82.
171. Calder PC. n-3 polyunsaturated fatty acids, inflammation, and inflammatory diseases. Am J Clin Nutr 2006; 83:1505S-19S.
172. Hart BL, Hart LA, Thigpen AP, Willits NH. Long-term health effects of neutering dogs: comparison of Labrador Retrievers with Golden Retrievers. PLoS One 2014; 9:e102241.
173. Zink MC, Farhoody P, Elser SE, Ruffini LD, Gibbons TA, Rieger RH. Evaluation of the risk and age of onset of cancer and behavioral disorders in gonadectomized Vizslas. J Am Vet Med Assoc 2014; 244:309-19.
174. Hoffman JM, Creevy KE, Promislow DE. Reproductive capability is associated with lifespan and cause of death in companion dogs. PLoS One 2013; 8:e61082.
175. caret: Classification and Regression Training. R package version 6.0-73. 2016. Available from https://cran.r-project.org/package=caret.
176. Kleinbaum DG, Klein M. Logistic Regression: A Self-Learning Text. 3 ed. New York: Springer-Verlag; 2010.
177. Calder PC. Polyunsaturated fatty acids, inflammation, and immunity. Lipids 2001; 36:1007-24.
178. Calder PC. Sir David Cuthbertson Medal Lecture. Immunomodulatory and anti-inflammatory effects of n-3 polyunsaturated fatty acids. Proc Nutr Soc 1996; 55:737-74.

179. Kinsella JE, Lokesh B, Broughton S, Whelan J. Dietary polyunsaturated fatty acids and eicosanoids: potential effects on the modulation of inflammatory and immune cells: an overview. Nutrition 1990; 6:24-44; discussion 59-62.
180. Gosselin SJ, Capen CC, Martin SL. Histologic and ultrastructural evaluation of thyroid lesions associated with hypothyroidism in dogs. Vet Pathol 1981; 18:299-309.
181. Norris CR, Williams BJ, Ling GV, Franti CE, Johnson, Ruby AL. Recurrent and persistent urinary tract infections in dogs: 383 cases (1969-1995). J Am Anim Hosp Assoc 2000; 36:484-92.
182. Ling GV, Norris CR, Franti CE, Eisele PH, Johnson DL, Ruby AL, et al. Interrelations of organism prevalence, specimen collection method, and host age, sex, and breed among 8,354 canine urinary tract infections (1969-1995). J Vet Intern Med 2001; 15:341-7.
183. Ling GV. Therapeutic strategies involving antimicrobial treatment of the canine urinary tract. J Am Vet Med Assoc 1984; 185:1162-4.
184. McGhie JA, Stayt J, Hosgood GL. Prevalence of bacteriuria in dogs without clinical signs of urinary tract infection presenting for elective surgical procedures. Aust Vet J 2014; 92:33-7.
185. Wan SY, Hartmann FA, Jooss MK, Viviano KR. Prevalence and clinical outcome of subclinical bacteriuria in female dogs. J Am Vet Med Assoc 2014; 245:106-12.
186. Anderson LA, Atman AA, McShane CM, Titmarsh GJ, Engels EA, Koshiol J. Common infection-related conditions and risk of lymphoid malignancies in older individuals. Br J Cancer 2014; 110:2796-803.
187. Brložnik M, Šterk K, Zdovc I. Prevalence and resistance patterns of canine uropathogens in regard to concurrent diseases. Berl Munch Tierarztl Wochenschr 2016; 129:340-50.
188. Engels EA, Cerhan JR, Linet MS, Cozen W, Colt JS, Davis S, et al. Immune-related conditions and immune-modulating medications as risk factors for non-Hodgkin's lymphoma: a case-control study. Am J Epidemiol 2005; 162:1153-61.
189. Hussell T, Isaacson PG, Crabtree JE, Spencer J. Helicobacter pylori-specific tumour-infiltrating T cells provide contact dependent help for the growth of malignant B cells in low-grade gastric lymphoma of mucosa-associated lymphoid tissue. J Pathol 1996; 178:122-7.
190. Hussell T, Isaacson PG, Crabtree JE, Spencer J. The response of cells from low-grade B-cell gastric lymphomas of mucosa-associated lymphoid tissue to Helicobacter pylori. Lancet 1993; 342:571-4.
191. Lecuit M, Abachin E, Martin A, Poyart C, Pochart P, Suarez F, et al. Immunoproliferative small intestinal disease associated with Campylobacter jejuni. N Engl J Med 2004; 350:239-48.
192. Gelatt KN, Gilger BC, Kern TJ. Veterinary Ophtalmology. 5th ed: John Wiley & Sons, Inc.; 2013.
193. Slatter's Fundamentals of Veterinary Ophthalmology. St. Louis, Missouri: Elsevier Inc.; 2008.
194. Donzel E, Arti L, Chahory S. Epidemiology and clinical presentation of canine cataracts in France: a retrospective study of 404 cases. Vet Ophthalmol 2017; 20:131-9.

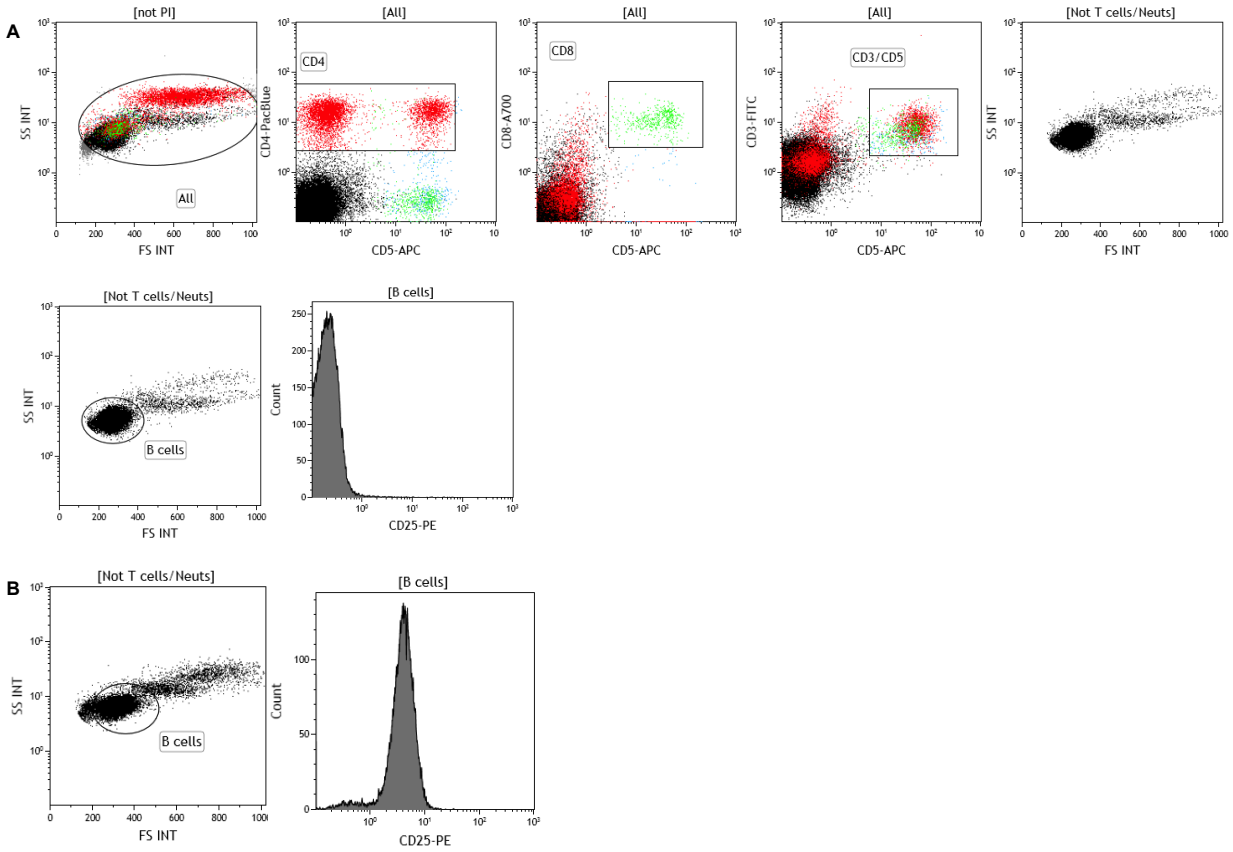
195. Chang ET, Bälter KM, Torrång A, Smedby KE, Melbye M, Sundström C, et al. Nutrient intake and risk of non-Hodgkin's lymphoma. Am J Epidemiol 2006; 164:1222-32.
196. Fritschi L, Ambrosini GL, Kliewer EV, Johnson KC, Group CCRER. Dietary fish intake and risk of leukaemia, multiple myeloma, and non-Hodgkin lymphoma. Cancer Epidemiol Biomarkers Prev 2004; 13:532-7.
197. Johansson AS, Norén-Nyström U, Larefalk A, Holmberg D, Lindskog M. Fish oil delays lymphoma progression in the TLL mouse. Leuk Lymphoma 2010; 51:2092-7.
198. Fallah M, Liu X, Ji J, Försti A, Sundquist K, Hemminki K. Autoimmune diseases associated with non-Hodgkin lymphoma: a nationwide cohort study. Ann Oncol 2014; 25:2025-30.
199. Mizokami T, Hamada K, Maruta T, Higashi K, Yamashita H, Noguchi Y, et al. Development of Primary Thyroid Lymphoma during an Ultrasonographic Follow-up of Hashimoto's Thyroiditis: A Report of 9 Cases. Intern Med 2016; 55:943-8.
200. Doody MM, Linet MS, Glass AG, Friedman GD, Pottern LM, Boice JD, et al. Leukemia, lymphoma, and multiple myeloma following selected medical conditions. Cancer Causes Control 1992; 3:449-56.
201. Legendre L, Barnetche T, Mazereeuw-Hautier J, Meyer N, Murrell D, Paul C. Risk of lymphoma in patients with atopic dermatitis and the role of topical treatment: A systematic review and meta-analysis. J Am Acad Dermatol 2015; 72:992-1002.
202. Smith MA, Simon R, Strickler HD, McQuillan G, Ries LA, Linet MS. Evidence that childhood acute lymphoblastic leukemia is associated with an infectious agent linked to hygiene conditions. Cancer Causes Control 1998; 9:285-98.
203. Hallek M. Chronic lymphocytic leukemia: 2015 Update on diagnosis, risk stratification, and treatment. Am J Hematol 2015; 90:446-60.
204. Molica S. Sex differences in incidence and outcome of chronic lymphocytic leukemia patients. Leuk Lymphoma 2006; 47:1477-80.
205. Morton LM, Wang SS, Devesa SS, Hartge P, Weisenburger DD, Linet MS. Lymphoma incidence patterns by WHO subtype in the United States, 1992-2001. Blood 2006; 107:265-76.
206. Wang M, Kaufman RJ. The impact of the endoplasmic reticulum protein-folding environment on cancer development. Nat Rev Cancer 2014; 14:581-97.
207. Sun S, Shi G, Han X, Francisco AB, Ji Y, Mendonça N, et al. Sel1L is indispensable for mammalian endoplasmic reticulum-associated degradation, endoplasmic reticulum homeostasis, and survival. Proc Natl Acad Sci U S A 2014; 111:E582-91.
208. Wen Z, Fiocchi C. Inflammatory bowel disease: autoimmune or immune-mediated pathogenesis? Clin Dev Immunol 2004; 11:195-204.
209. Larsen PR, Silva JE, Kaplan MM. Relationships between circulating and intracellular thyroid hormones: physiological and clinical implications. Endocr Rev 1981; 2:87-102.
210. Karube K, Scarfò L, Campo E, Ghia P. Monoclonal B cell lymphocytosis and "in situ" lymphoma. Semin Cancer Biol 2014; 24:3-14.
211. Landgren O, Kyle RA, Pfeiffer RM, Katzmann JA, Caporaso NE, Hayes RB, et al. Monoclonal gammopathy of undetermined significance (MGUS) consistently precedes multiple myeloma: a prospective study. Blood 2009; 113:5412-7.

212. Landgren O, Albitar M, Ma W, Abbasi F, Hayes RB, Ghia P, et al. B-cell clones as early markers for chronic lymphocytic leukemia. N Engl J Med 2009; 360:659-67.
213. Rawstron AC, Bennett FL, O'Connor SJ, Kwok M, Fenton JA, Plummer M, et al. Monoclonal B-cell lymphocytosis and chronic lymphocytic leukemia. N Engl J Med 2008; 359:575-83.
214. Crowther-Swanepoel D, Corre T, Lloyd A, Gaidano G, Olver B, Bennett FL, et al. Inherited genetic susceptibility to monoclonal B-cell lymphocytosis. Blood 2010; 116:5957-60.
215. Botstein D, Risch N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. Nat Genet 2003; 33 Suppl:228-37.
216. Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? J Med Genet 2011; 48:580-9.
217. El-Metwally S, Hamza T, Zakaria M, Helmy M. Next-generation sequence assembly: four stages of data processing and computational challenges. PLoS Comput Biol 2013; 9:e1003345.
218. Abnizova I, te Boekhorst R, Orlov YL. Computational Errors and Biases in Short Read Next Generation Sequencing. J Proteomics
Bioinform 2017; 10:1-17.
219. Botta V, Louppe G, Geurts P, Wehenkel L. Exploiting SNP correlations within random forest for genome-wide association studies. PLoS One 2014; 9:e93379.
220. Thornton-Wells TA, Moore JH, Haines JL. Genetics, statistics and human disease: analytical retooling for complexity. Trends Genet 2004; 20:640-7.
221. Sebastiani P, Timofeev N, Dworkis DA, Perls TT, Steinberg MH. Genome-wide association studies and the genetic dissection of complex traits. Am J Hematol 2009; 84:504-15.
222. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. Nat Rev Genet 2009; 10:392-404.
223. Yoo W, Ference BA, Cote ML, Schwartz A. A Comparison of Logistic Regression, Logic Regression, Classification Tree, and Random Forests to Identify Effective Gene-Gene and Gene-Environmental Interactions. Int J Appl Sci Technol 2012; 2:268.
224. Oh S, Lee J, Kwon MS, Weir B, Ha K, Park T. A novel method to identify high order gene-gene interactions in genome-wide association studies: gene-based MDR. BMC Bioinformatics 2012; 13 Suppl 9:S5.

APPENDIX 1: SUPPLEMENTARY FIGURES AND TABLES FOR AIM 1




**Suppl. Figure 2.1. Gating strategy for evaluation of class II MHC levels on B cells in peripheral blood**. Propidium iodide (PI) positive cells were first gated out to remove dead cells from the analysis. Then a gate was drawn around CD21 positive cells (blue, left column). Class II MHC levels were evaluated on CD21 positive cells (histogram, right column). The grey cells in the left column consist of neutrophils, monocytes and T cells, as determined by staining reactions in other tubes. (A) 11 year old MN Maltese with a CD21 count of 60,600 cells/μL, exhibiting high levels of class II MHC. (B) 12- year old FS Bichon Frise with a CD21 count of 8,000 cells/μL, exhibiting intermediate levels of class II MHC. (C) 6 year old MN English Bulldog with a CD21 count of 55,412 cells/μL, exhibiting low levels of class II MHC.




**Suppl. Figure 2.2. Gating strategy to detect CD25 expression on B cells.** Anti-CD25 and anti-CD21 were not in the same staining reaction, so the strategy shown above was used to measure CD25 expression on B cells. First, a gate was drawn around all peripheral blood cells. Then gates were drawn around CD4 expressing cells (T cells and neutrophils), CD8 expressing cells, and any cells expressing CD3 and CD5 (to eliminate CD4-CD8- T cells). These three gates were combined to create a single "not" Boolean gate, which eliminates these cells. The panel on the far right of the top row shoes that this gating strategy leaves primarily small lymphocytes, which is where the CD21 expressing cells are concentrated. A gate is then drawn around these small cells (B cells), and CD25 expression is evaluated (second row). (A) English Bulldog with B cells that do not express CD25. (B) Mixed breed dog in which all the B cells express CD25.

**Suppl. Table 2.1. Detailed CBC findings for 128 anemic dogs with B-CLL.**

| | | Dogs with available data (%) | n | % |
|---|---|---|---|---|
| **Neutrophil Total** | Neutropenia (< 2,000/ul) | 99.2 | 2 | 1.6 |
| | Normal (2,000 - 12,000/ul) | | 88 | 69.3 |
| | Neutrophilia (> 12,000/ul) | | 37 | 29.1 |
| **Platelet Total** | Thrombocytopenia (<175,000/ul) | 71.9 | 14 | 15.2 |
| | Normal (175,000-500,000/ul) | | 63 | 68.5 |
| | Thrombocytosis (>500,000/ul) | | 15 | 16.3 |
| **Regenerative Anemia?** | Yes (Reticulocytes >100,000/ul) | 52.3 | 21 | 31.3 |
| | No (Reticulocytes <100,000/ul) | | 46 | 68.7 |
| **Hematocrit** | *Median (IQR)* | 100 | 30 | (26 - 34) |

**Suppl. Table 2.2. Number of cases, total number of submissions, and percent of submissions diagnosed as B-CLL for each breed.** Odds ratios and 95% confidence intervals comparing B-CLL among each breed are compared using mixed breed dogs and Labrador Retrievers as a reference. Only breeds with at least 30 overall submissions were considered. Bolded values are statistically significant ($p<0.05$).

| | | | | vs Mixed Breed | | vs Labrador Retriever | |
|---|---|---|---|---|---|---|---|
| **Breed** | **CLL** | **Total** | **%CLL** | **OR** | **95% CI** | **OR** | **95% CI** |
| Dachshund | 14 | 49 | 29% | **4.48** | **(2.34 - 8.57)** | **8.10** | **(3.84 - 17.12)** |
| Pomeranian | 9 | 34 | 26% | **4.03** | **(1.84 - 8.84)** | **7.29** | **(3.06 - 17.40)** |
| Cairn Terrier | 8 | 31 | 26% | **3.89** | **(1.70 - 8.90)** | **7.05** | **(2.85 - 17.45)** |
| Bichon Frise | 11 | 46 | 24% | **3.52** | **(1.74 - 7.12)** | **6.37** | **(2.87 - 14.12)** |
| English Bulldog | 23 | 99 | 23% | **3.39** | **(2.04 - 5.61)** | **6.13** | **(3.28 - 11.48)** |
| Boston Terrier | 7 | 34 | 21% | **2.90** | **(1.24 - 6.81)** | **5.25** | **(2.07 - 13.32)** |
| Jack Russell Terrier | 14 | 69 | 20% | **2.85** | **(1.54 - 5.28)** | **5.16** | **(2.51 - 10.60)** |
| Maltese | 11 | 56 | 20% | **2.74** | **(1.38 - 5.44)** | **4.95** | **(2.27 - 10.81)** |
| Yorkshire Terrier | 11 | 61 | 18% | **2.46** | **(1.25 - 4.86)** | **4.46** | **(2.05 - 9.68)** |
| Shih Tzu | 28 | 177 | 16% | **2.10** | **(1.34 - 3.29)** | **3.81** | **(2.13 - 6.81)** |
| Cocker Spaniel | 14 | 91 | 15% | **2.03** | **(1.12 - 3.71)** | **3.68** | **(1.82 - 7.47)** |
| Chihuahua | 9 | 64 | 14% | 1.83 | (0.88 - 3.80) | **3.32** | **(1.46 - 7.53)** |
| Pit bull | 15 | 108 | 14% | **1.81** | **(1.01 - 3.22)** | **3.27** | **(1.64 - 6.50)** |
| West Highland White Terrier | 8 | 60 | 13% | 1.72 | (0.80 - 3.72) | **3.12** | **(1.33 - 7.32)** |
| Doberman | 5 | 39 | 13% | 1.65 | (0.63 - 4.29) | **2.98** | **(1.07 - 8.33)** |
| Boxer | 31 | 261 | 12% | 1.51 | (0.99 - 2.30) | **2.73** | **(1.56 - 4.79)** |
| Siberian Husky | 5 | 44 | 11% | 1.44 | (0.55 - 3.71) | 2.60 | (0.94 - 7.21) |
| Border Collie | 6 | 61 | 10% | 1.22 | (0.51 - 2.90) | 2.21 | (0.86 - 5.66) |
| Miniature Schnauzer | 5 | 52 | 10% | 1.19 | (0.46 - 3.05) | 2.16 | (0.78 - 5.93) |
| Shetland Sheepdog | 4 | 44 | 9% | 1.12 | (0.39 - 3.19) | 2.03 | (0.67 - 6.15) |
| Mixed Breed | 111 | 1353 | 8% | | Ref | 1.81 | (1.14 - 2.87) |
| Cavalier King Charles Spaniel | 3 | 39 | 8% | 0.93 | (0.28 - 3.08) | 1.69 | (0.48 - 5.89) |
| French Bulldog | 2 | 42 | 5% | 0.56 | (0.13 - 2.35) | 1.01 | (0.23 - 4.45) |
| Labrador Retriever | 23 | 489 | 5% | **0.55** | **(0.35 - 0.88)** | | Ref |
| Beagle | 6 | 140 | 4% | 0.50 | (0.22 - 1.16) | 0.91 | (0.36 - 2.27) |
| Australian Shepherd | 3 | 77 | 4% | 0.45 | (0.14 - 1.46) | 0.82 | (0.24 - 2.80) |
| Greyhound | 1 | 30 | 3% | 0.39 | (0.05 - 2.86) | 0.70 | (0.09 - 5.36) |
| Scottish Terrier | 1 | 32 | 3% | 0.36 | (0.05 - 2.67) | 0.65 | (0.09 - 5.00) |
| Mastiff | 1 | 37 | 3% | 0.31 | (0.04 - 2.29) | 0.56 | (0.07 - 4.29) |
| Rhodesian Ridgeback | 1 | 39 | 3% | 0.29 | (0.04 - 2.17) | 0.53 | (0.07 - 4.06) |
| Basset Hound | 1 | 62 | 2% | 0.18 | (0.03 - 1.34) | 0.33 | (0.04 - 2.50) |
| German Shepherd | 1 | 134 | 1% | **0.08** | **(0.01 - 0.61)** | 0.15 | (0.02 - 1.14) |
| Golden Retriever | 3 | 836 | 0% | **0.04** | **(0.01 - 0.13)** | **0.07** | **(0.02 - 0.24)** |
| Rottweiler | **0** | **85** | **0%** | -- | -- | -- | -- |
| Welsh Corgi | **0** | **75** | **0%** | -- | -- | -- | -- |
| Bernese Mountain Dog | **0** | **66** | **0%** | -- | -- | -- | -- |
| Standard Poodle | **0** | **50** | **0%** | -- | -- | -- | -- |

**Suppl. Table 2.3. Signalment, physical examination findings, and laboratory findings for 491 cases of B-CLL.** Findings are compared between Labrador Retrievers, mixed breed dogs, and all other breeds.

| | | | Dogs with available data (%) | Labrador Retrievers (n=23) | | Mixed Breeds (n=111) | | All Other Breeds (n=357) | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | n | % | n | % | n | % |
| **Signalment** | **Sex** | Male (intact or neutered) | 99.2 | 7 | 30.4 | 48 | 43.6 | 187 | 52.8 |
| | | Female (intact or neutered) | | 16 | 69.6 | 62 | 56.4 | 167 | 47.2 |
| | **Age,** *Median (IQR)* | | 98.4 | 11 | (9 - 13)* | 11 | (9 - 13) | 11.0 | (9 - 13) |
| **Veterinarian-reported Physical Exam Findings** | **Peripheral Lymphadenopathy** | | 62.7 | 11 | 73.3 | 37 | 52.3 | 93 | 41.7 |
| | **Splenomegaly** | | 46.0 | 4 | 40.0 | 26 | 54.2 | 85 | 50.6 |
| | **Hepatomegaly** | | 41.1 | 1 | 12.5 | 12 | 26.7 | 45 | 30.2 |
| | **Mediastinal Mass** | | 34.0 | 0 | 0.0 | 0 | 0.0 | 5 | 4.0 |
| | **Visceral Lymphadenopathy** | | 35.8 | 3 | 37.5 | 10 | 26.3 | 28 | 21.5 |
| **Veterinarian-reported Laboratory Findings** | **Hyperglobulinemia** | | 62.7 | 2 | 12.3 | 16 | 23.2 | 63 | 28.0 |
| | **Hypercalcemia** | | 58.9 | 0 | 0.0 | 2 | 3.2 | 11 | 5.2 |
| **CBC findings** | **Neutrophil Total** | Neutropenia (<2,000/ul) | | 0 | 0.0 | 1 | 1.0 | 4 | 1.1 |
| | | Normal (2,000 - 12,000/ul) | | 17 | 77.3 | 94 | 84.7 | 283 | 79.5 |
| | | Neutrophilia (>12,000/ul) | 99.6 | 5 | 22.7 | 16 | 14.4 | 69 | 19.4 |
| | **Platelet Total** | Thrombocytopenia (<175,000/ul and no clumps noted) | | 2 | 12.5 | 5 | 5.4 | 21 | 7.1 |
| | | Normal (175,000-500,000/ul) | | 13 | 81.3 | 77 | 82.8 | 220 | 74.6 |
| | | Thrombocytosis (>500,000/ul) | 82.3 | 1 | 6.3 | 11 | 11.8 | 54 | 18.3 |
| | **Anemia** | Hematocrit <36 | 99.4 | 0 | 0.0 | 28 | 25.5 | 99 | 27.9 |
| | **Lymphocyte Total**, *Median cells x $10^3$/ul (IQR)* | | 100 | 45.8 | (17.1 - 69.5) | 21.9 | (14.3 - 52.9) | 24.6 | (13.7 - 47.0) |

*: values for age and lymphocyte total are given as median (IQR)

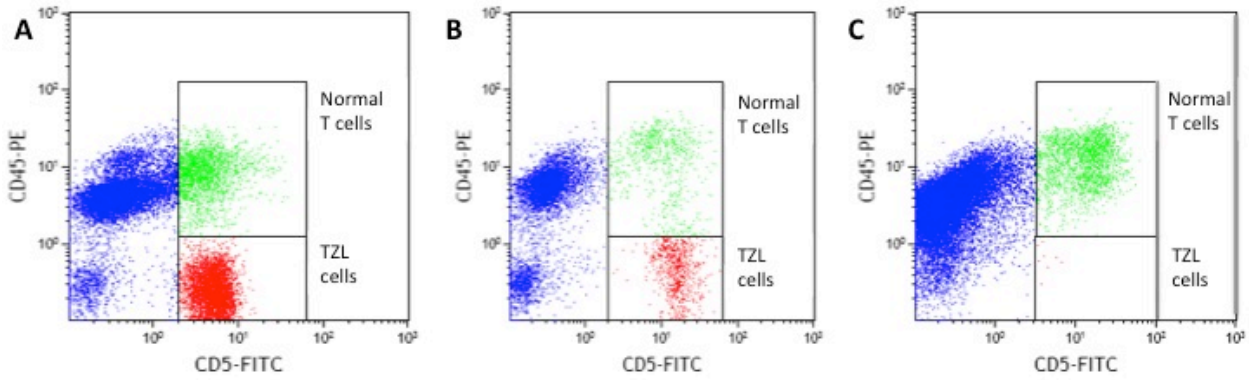
APPENDIX 2: SUPPLEMENTARY FIGURES AND TABLES FOR AIM 2

**Suppl. Table 3.1. Antibody panels used for immunophenotyping.**

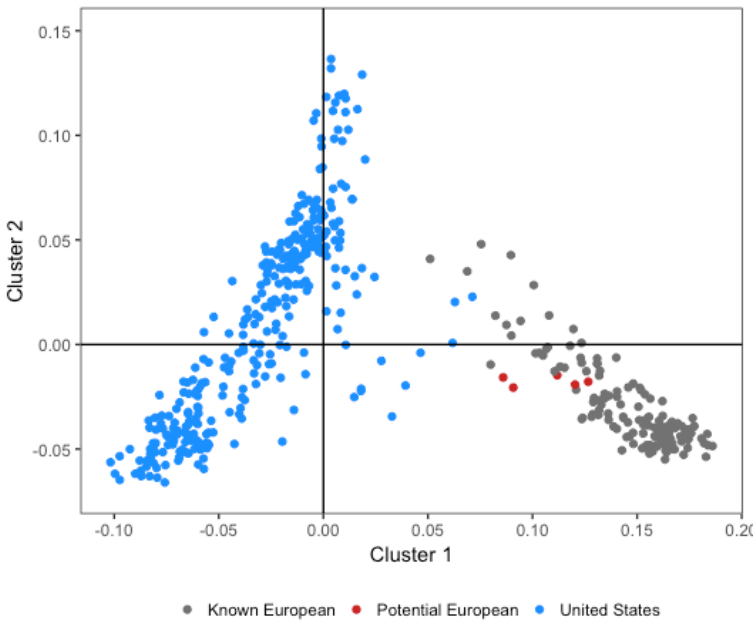
| Tube | Antibody specificity and fluorochrome |
|---|---|
| 1 | M IgG1-FITC / M IgG1-PE / M IgG1-Alexa 647 / M IgG1-Alexa 700 / M IgG1-PE-750 / M IgG1-Pacific Blue |
| 2 | CD3-FITC / CD25-PE / CD5-APC / CD8-Alexa 700 / CD4-Pacific Blue |
| 3 | Class II MHC-FITC / CD22-PE / CD21-Alexa 647 |
| 4 | Class II MHC-FITC / CD34-PE / CD5-APC - CD14-PE-Alexa 750 |
| 5 | Class II MHC-FITC / CD18-PE / CD5-APC / CD14 PE-Alexa 750 / CD4-Pacific Blue |
| 6 | CD5-FITC / CD45-PE / CD21-Alexa 647 |

***M= mouse**

Unless otherwise noted, all antibodies were purchased from AbD Serotec. Clones are as follows: CD45 = YKIX716.13, CD18 = YFC118.3 (human CD18), CD4 = YKIX302.9, CD8 = YCATE 55.9, CD5 = YKIX322.3, CD21 = CA2.1D6, CD22 = RFB4 (human CD22, purchased from AbCam), CD3 = CA17.2A12, CD14 = UCHM (human, used in panel 1) and CD14 = TUK4 (human, used in panel 2), class II MHC = YKIX334.2, CD34 = 1H6, CD25 = P2A10 (purchased from eBiosciences)



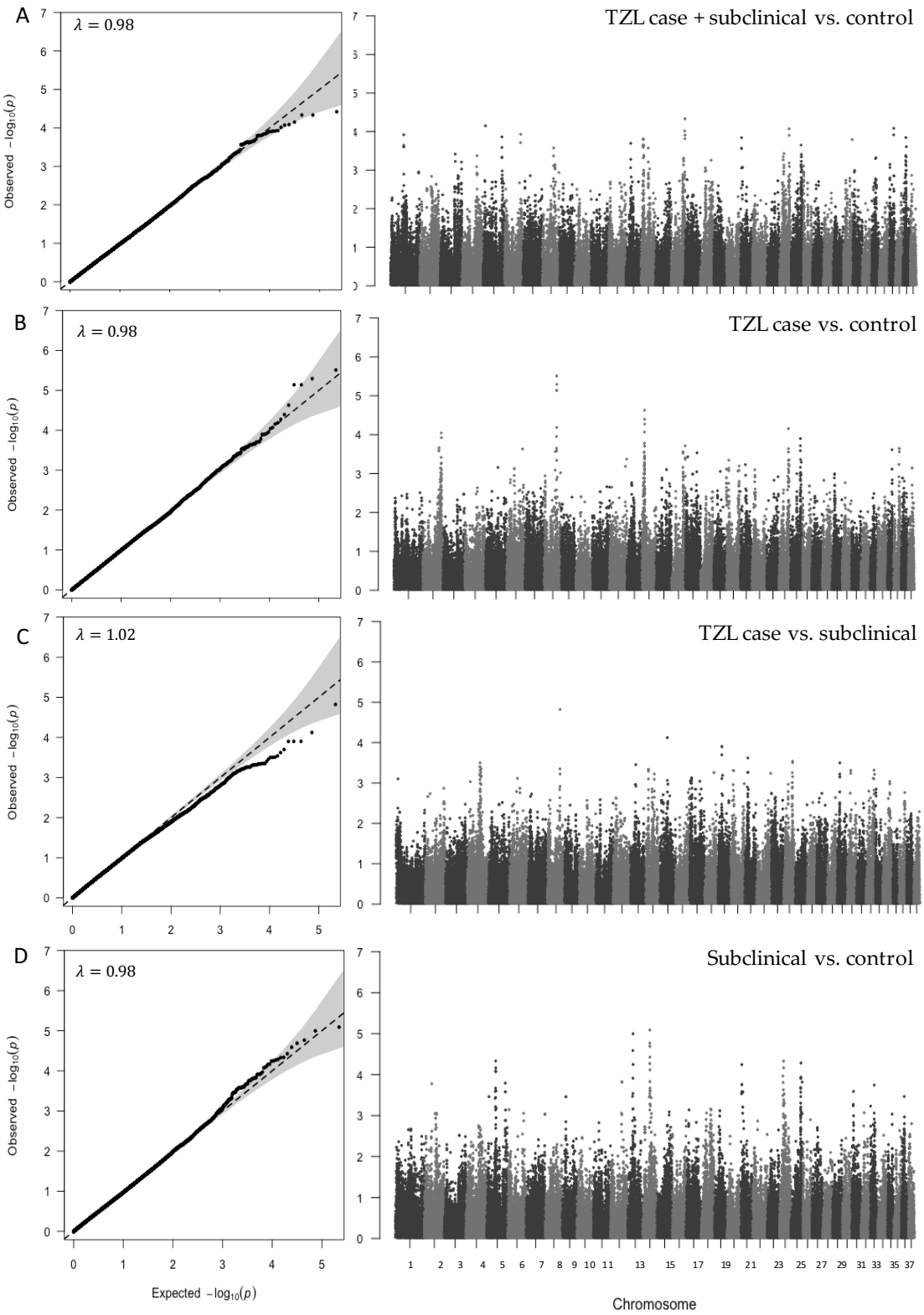

**Suppl. Fig 3.1. Flow cytometric analysis of peripheral blood samples**. (A) Sample considered diagnostic for TZL due to homogeneous expansion of $CD5^{+}CD45^{-}$ T cells (red cells). (B) Sample diagnosed as subclinical due to smaller population of $CD5^{+}CD45^{-}$ T cells and absence of lymphocytosis or lymphadenopathy. (C) Sample considered a control; all T cells are CD5+CD45+ (green cells; normal)



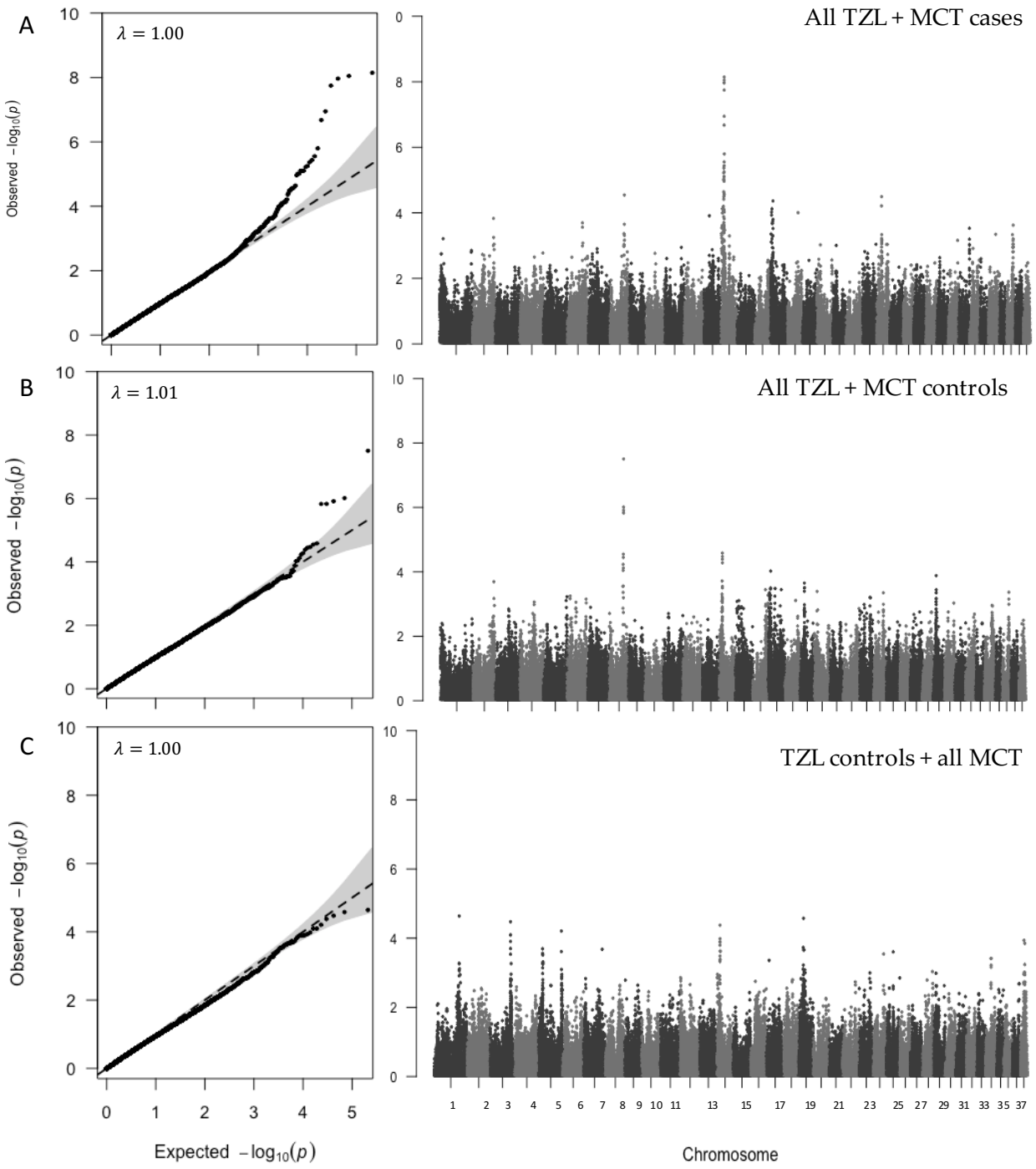

**Suppl. Fig 3.2. Multidimensional scaling plot showing clustering of European dogs**. Dogs marked in red met our threshold for European origin based on clustering and were subsequently removed from the analysis.

**Suppl. Table 3.2. Signalment, GWAS haplotype, and coverage for 16 resequenced dogs.**

| | | | | | | Chromosome 8 | | | Chromosome 14 | |
|---|---|---|---|---|---|---|---|---|---|---|
| Run | ID | Sex | Age (years) | State | Phenotype | GWAS haplotype | Resequencing coverage [median (IQR)] | Hypothyroid? | GWAS haplotype | Resequencing coverage [median (IQR)] |
| 1 | MAF029 | MI | 18 | CO | Subclinical | nn | 129 (14 - 160) | No | nn | 135 (98 - 159) |
| 3 | MAF106 | MC | 15 | CO | Control | nn | 125 (15 - 153) | Yes | rn | 133 (99 - 156) |
| 1 | MAF146 | FS | 11 | IL | Control | nn | 114 (14 - 141) | Yes | rn | 121 (88 - 144) |
| 4 | RFCo038 | MC | 11 | OK | Subclinical | nn | 141 (20 - 170) | No | rn | 149 (111 - 172) |
| 2 | RFCa066 | FS | 10 | MA | Case | nn | 136 (21 - 165) | No | rr | 143 (108 - 166) |
| 3 | RFCa110 | FS | 9 | NJ | Case | rn | 128 (16 - 157) | No | rr | 137 (101 - 159) |
| 2 | RFCa115 | MC | 12 | WA | Case | nn | 131 (19 - 158) | No | rr | 154 (118 - 177) |
| 3 | RFCa136 | MC | 11 | VA | Case | rn | 133 (17 - 162) | No | rn | 139 (103 - 162) |
| 2 | MAF285 | MC | 10 | TX | Subclinical | nn | 135 (20 - 164) | No | nn | 142 (107 - 164) |
| 2 | RFCa093 | FS | 12 | TX | Case | rn | 134 (20 - 163) | No | rr | 179 (141 - 201) |
| 1 | MAF109 | FS | 7 | NV | Case | nn | 115 (12 - 143) | No | rr | 123 (89 - 146) |
| 4 | MAF290 | MC | 10 | UT | Subclinical | rr | 137 (18 - 166) | Yes | n-recomb | 145 (107 - 168) |
| 1 | MAF301 | MC | 9 | AR | Control | nn | 127 (15 - 156) | Yes | rn | 134 (99 - 158) |
| 3 | RFCa143 | MC | 12 | TN | Case | rn | 107 (13 - 133) | No | rn | 115 (84 - 136) |
| 4 | RFCa085 | FS | 13 | OK | Case | rn | 132 (17 - 161) | No | rr | 141 (104 - 164) |
| 4 | RFCo105 | MC | 9 | GA | Subclinical | nn | 140 (20 - 168) | No | nn | 148 (111 - 171) |




**Suppl. Fig. 3.3. GWAS results for all TZL combinations.** QQ plot (left) and Manhattan plot (right). (A) TZL case + subclinical vs. control. (B) TZL case vs. control. (C) TZL case vs. subclinical. (D) Subclinical vs. control.




**Suppl. Fig. 3.4. GWAS results for combined TZL and MCT datasets.** QQ plot (left) and Manhattan plot (right). (A) All TZL + MCT cases (B) All TZL + MCT controls. (C) TZL subclinical + TZL control + all MCT.

**Suppl. Table 3.3. Summary of SnpEff annotations for single nucleotide variants.**

| GeneName | BioType | Approximate gene location | Moderate impact | Missense | Modifier | 5' UTR | 3' UTR | Upstream gene | Downstream gene | Intron | Non-coding transcript exon | Low impact | Splice region | Synonymous |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Chromosome 8** | | | | | | | | | | | | | | |
| ENSCAFG00000027706 | snoRNA | 52,401,055-52,401,186 | - | - | 7 | - | - | 5 | 2 | - | - | - | - | - |
| ENSCAFG00000035649 | lincRNA | 52,843,084-52,872,396 | - | - | 11 | - | - | - | - | 11 | - | - | - | - |
| CEP128 | Protein coding | 52,914,417-53,290,253 | - | - | 170 | - | 3 | - | 11 | 156 | - | - | - | - |
| ENSCAFG00000039638 | lincRNA | 53,309,256-53,312,473 | - | - | 1 | - | - | - | - | 1 | - | - | - | - |
| GTF2A1 | Protein coding | 53,519,940-53,564,001 | - | - | 36 | - | 2 | 5 | 2 | 27 | - | - | - | - |
| ENSCAFG00000026835 | snoRNA | 53,551,375-53,551,519 | - | - | 6 | - | - | 2 | 4 | - | - | - | - | - |
| ENSCAFG00000037858 | lincRNA | 53,566,146-53,579,803 | - | - | 9 | - | - | 4 | 1 | 4 | - | - | - | - |
| STON2 | Protein coding | 53,597,941-53,725,688 | - | - | 127 | - | - | 1 | - | 126 | - | - | - | - |
| ENSCAFG00000033948 | lincRNA | 53,746,557-53,759,679 | - | - | 21 | - | - | 4 | 11 | 6 | - | - | - | - |
| SEL1L | Protein coding | 53,771,761-53,824,492 | - | - | 211 | - | - | 16 | 9 | 186 | - | 3 | - | 3 |
| ENSCAFG00000039987 | lincRNA | 53,805,971-53,840,850 | - | - | 119 | - | - | 23 | 2 | 93 | 1 | - | - | - |
| **Chromosome 14** | | | | | | | | | | | | | | |
| IMPDH1 | Protein coding | 7,996,893-8,013,287 | - | - | 1 | - | - | - | - | 1 | - | - | - | - |
| PRRT4 | Protein coding | 8,029,236-8,038,095 | - | - | 1 | - | - | - | - | 1 | - | - | - | - |
| LEP | Protein coding | 8,114,472-8,131,160 | - | - | 1 | - | - | - | 1 | - | - | - | - | - |
| ENSCAFG00000036574 | lincRNA | 8,141,376-8,144,851 | - | - | 1 | - | - | - | 1 | - | - | - | - | - |
| ENSCAFG00000032858 | lincRNA | 8,148,588-8,152,852 | - | - | 1 | - | - | 1 | - | - | - | - | - | - |
| ENSCAFG00000034487 | lincRNA | 8,212,887-8,226,291 | - | - | 1 | - | - | 1 | - | - | - | - | - | - |
| SND1 | Protein coding | 8,260,919-8,685,952 | - | - | 11 | - | - | 1 | - | 10 | - | - | - | - |
| PAX4 | Protein coding | 8,714,949-8,719,371 | - | - | 8 | - | - | 4 | 4 | 3 | - | - | - | - |
| FSCN3 | Protein coding | 8,724,865-8,730,729 | - | - | 11 | - | - | 1 | 6 | 8 | - | - | - | - |
| ARF5 | Protein coding | 8,728,365-8,735,331 | - | - | 13 | - | 4 | 3 | 6 | - | - | - | - | - |
| GCC1 | Protein coding | 8,738,782-8,742,022 | - | - | 3 | - | - | 3 | - | - | - | - | - | - |
| ENSCAFG00000035664 | lincRNA | 8,779,984-8,787,289 | - | - | 4 | - | - | - | 2 | 1 | 1 | - | - | - |
| ZNF800 | Protein coding | 8,866,900-8,884,790 | - | - | 3 | - | - | 2 | - | 1 | - | - | - | - |
| GRM8 | Protein coding | 8,993,463-9,719,899 | - | - | 248 | - | - | 1 | 2 | 245 | - | 1 | - | 1 |
| cif-mir-592 | miRNA | 9,182,237-9,182,331 | - | - | 5 | - | - | 3 | 2 | - | - | - | - | - |
| ENSCAFG00000037039 | lincRNA | 9,988,593-9,989,896 | - | - | 1 | - | - | - | 1 | - | - | - | - | - |
| ENSCAFG00000036162 | lincRNA | 10,190,965-10,354,356 | - | - | 86 | - | - | - | - | 86 | - | - | - | - |
| POT1 | Protein coding | 10,986,871-11,070,651 | - | - | 4 | - | - | - | - | 4 | - | - | - | - |
| GPR37 | Protein coding | 11,114,415-11,132,395 | - | - | 4 | - | - | 2 | - | 2 | - | - | - | - |
| ENSCAFG00000001759 | Protein coding | 11,350,342-11,351,840 | - | - | 18 | - | - | 7 | 11 | - | - | 3 | - | 3 |
| ENSCAFG00000032124 | Processed pseudogene | 11,447,820-11,448,038 | - | - | 13 | - | - | 10 | 3 | - | - | - | - | - |
| LSM3 | Protein coding | 11,481,430-11,481,735 | - | - | 1 | - | - | - | 1 | - | - | - | - | - |
| TMEM229A | Protein coding | 11,646,915-11,648,024 | - | - | 5 | - | - | - | 5 | - | - | - | - | - |
| SPAM1 | Protein coding | 11,703,906-11,723,178 | 1 | 1 | 51 | - | 3 | 5 | 7 | 36 | - | - | - | - |
| HYAL4 | Protein coding | 11,736,206-11,743,786 | 3 | 3 | 58 | 2 | 1 | 21 | 25 | 11 | - | - | - | - |
| ENSCAFG00000024436 | Protein coding | 11,760,627-11,770,158 | 1 | 1 | 27 | - | - | 7 | 12 | 8 | - | 1 | - | 1 |
| ENSCAFG00000026099 | snoRNA | 11,780,392-11,780,603 | - | - | 15 | - | - | 11 | 3 | - | 1 | - | - | - |
| WASL | Protein coding | 11,828,687-11,901,682 | - | - | 102 | - | - | 13 | 1 | 88 | - | 1 | - | 1 |
| LMOD2 | Protein coding | 11,918,483-11,925,130 | - | - | 23 | - | - | 6 | 7 | 10 | - | - | - | - |
| ASB15 | Protein coding | 11,931,570-11,958,359 | - | - | 23 | - | - | 3 | 6 | 14 | - | - | - | - |
| ENSCAFG00000025112 | Protein coding | 11,995,994-12,008,054 | - | - | 4 | - | - | 4 | - | - | - | - | - | - |
| GRM3 | Protein coding | 12,980,180-13,071,783 | - | - | 55 | - | - | - | - | 55 | - | 1 | - | 1 |
| KIAA1324L | Protein coding | 13,076,932-13,245,320 | - | - | 16 | - | - | - | - | 16 | - | - | - | - |
| DMTF1 | Protein coding | 13,338,261-13,377,136 | - | - | 25 | - | - | 4 | 1 | 20 | - | 1 | - | 1 |
| TMEM243 | Protein coding | 13,378,676-13,381,317 | - | - | 4 | - | - | 1 | 2 | 1 | - | - | - | - |
| ENSCAFG00000039543 | lincRNA | 13,424,493-13,432,523 | - | - | 2 | - | - | 1 | 1 | - | - | - | - | - |
| CROT | Protein coding | 13,474,854-13,534,793 | 1 | 1 | 173 | - | 9 | 4 | 21 | 150 | - | 2 | - | 2 |
| ABCB4 | Protein coding | 13,542,668-13,614,146 | - | - | 36 | - | - | 3 | 3 | 31 | - | 1 | 1 | - |
| MDR1 | Protein coding | 13,644,042-13,743,053 | 2 | 2 | 138 | - | - | - | - | 138 | - | - | - | - |
| RUNDC3B | Protein coding | 13,803,600-13,930,733 | - | - | 4 | - | - | 1 | - | 3 | - | - | - | - |
| DBF4 | Protein coding | 13,983,320-14,019,701 | - | - | 5 | - | - | - | 1 | 4 | - | - | - | - |
| ENSCAFG00000038610 | lincRNA | 13,984,440-13,996,438 | - | - | 2 | - | - | - | - | 2 | - | - | - | - |
| ADAM22 | Protein coding | 14,032,475-14,250,288 | - | - | 3 | - | - | - | - | 3 | - | 1 | - | 1 |
| SRI | Protein coding | 14,260,113-14,275,207 | - | - | 1 | - | - | - | - | 1 | - | - | - | - |
| STEAP4 | Protein coding | 14,319,840-14,345,870 | - | - | 23 | - | - | 22 | - | 1 | - | - | - | - |
| ZNF804B | Protein coding | 15,146,202-15,247,064 | - | - | 1 | - | - | - | - | 1 | - | - | - | - |
| ENSCAFG00000037438 | lincRNA | 15,322,541-15,348,896 | - | - | 93 | - | - | - | 9 | 84 | - | - | - | - |
| ENSCAFG00000038636 | lincRNA | 15,404,763-15,519,171 | - | - | 19 | - | - | - | 2 | 17 | - | - | - | - |
| ENSCAFG00000021475 | snRNA | 15,626,361-15,626,464 | - | - | 23 | - | - | 13 | 10 | - | - | - | - | - |
| ENSCAFG00000040353 | lincRNA | 15,711,267-15,714,286 | - | - | 9 | - | - | 6 | 3 | - | - | - | - | - |
| ENSCAFG00000034336 | lincRNA | 15,948,486-15,996,712 | - | - | 2 | - | - | - | - | 2 | - | - | - | - |
| STEAP1 | Protein coding | 16,074,785-16,088,182 | - | - | 3 | - | - | 2 | - | 1 | - | - | - | - |

**Suppl. Table 3.4. Summary of SnpEff annotations for indels.**

| GeneName | BioType | Approximate gene location | Moderate impact | Disruptive inframe deletion | Conservative inframe deletion | Modifier | 5' UTR | 3' UTR | Upstream gene | Downstream gene | Intron |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Chromosome 8** | | | | | | | | | | | |
| ENSCAFG00000035649 | lincRNA | 52,843,084-52,872,396 | - | - | - | 6 | - | - | - | - | 6 |
| CEP128 | Protein coding | 52,914,417-53,290,253 | - | - | - | 52 | - | 1 | - | 8 | 43 |
| TSHR | Protein coding | 53,327,081-53,483,111 | - | - | - | 2 | - | - | - | - | 2 |
| GTF2A1 | Protein coding | 53,519,940-53,564,001 | - | - | - | 16 | - | - | - | 2 | 14 |
| ENSCAFG00000026835 | snoRNA | 53,551,375-53,551,519 | - | - | - | 3 | - | - | - | 3 | - |
| ENSCAFG00000037858 | lincRNA | 53,566,146-53,579,803 | - | - | - | 5 | - | - | 1 | 1 | 3 |
| STON2 | Protein coding | 53,597,941-53,725,688 | - | - | - | 29 | - | - | - | - | 29 |
| ENSCAFG00000033948 | lincRNA | 53,746,557-53,759,679 | - | - | - | 2 | - | - | 1 | 1 | - |
| SEL1L | Protein coding | 53,771,761-53,824,492 | - | - | - | 68 | - | - | 6 | 5 | 57 |
| ENSCAFG00000039987 | lincRNA | 53,805,971-53,840,850 | - | - | - | 30 | - | - | 10 | - | 20 |
| **Chromosome 14** | | | | | | | | | | | |
| ENSCAFG00000001744 | Protein coding | 10,151,131-10,151,999 | - | - | - | 1 | - | - | - | 1 | - |
| ENSCAFG00000036162 | lincRNA | 10,190,965-10,354,356 | - | - | - | 16 | - | - | - | - | 16 |
| POT1 | Protein coding | 10,986,871-11,070,651 | - | - | - | 6 | - | - | - | - | 6 |
| GPR37 | Protein coding | 11,114,415-11,132,395 | - | - | - | 2 | - | - | 1 | - | 1 |
| ENSCAFG00000001759 | Protein coding | 11,350,342-11,351,840 | - | - | - | 3 | - | - | - | 3 | - |
| ENSCAFG00000032124 | Processed pseudogene | 11,447,820-11,448,038 | - | - | - | 1 | - | - | 1 | - | - |
| SPAM1 | Protein coding | 11,703,906-11,723,178 | - | - | - | 10 | - | - | 2 | 1 | 7 |
| HYAL4 | Protein coding | 11,736,206-11,743,786 | - | - | - | 18 | 1 | 2 | 5 | 7 | 4 |
| ENSCAFG00000024436 | Protein coding | 11,760,627-11,770,158 | - | - | - | 13 | - | - | 1 | 4 | 8 |
| ENSCAFG00000026099 | snoRNA | 11,780,392-11,780,603 | - | - | - | 4 | - | - | 2 | 2 | - |
| WASL | Protein coding | 11,828,687-11,901,682 | - | - | - | 52 | - | - | 9 | - | 43 |
| LMOD2 | Protein coding | 11,918,483-11,925,130 | 2 | 1 | 1 | 4 | - | - | 2 | - | 2 |
| ASB15 | Protein coding | 11,931,570-11,958,359 | - | - | - | 7 | - | - | 1 | 3 | 3 |
| GRM3 | Protein coding | 12,980,180-13,071,783 | - | - | - | 21 | - | - | - | - | 21 |
| KIAA1324L | Protein coding | 13,076,932-13,245,320 | - | - | - | 3 | - | - | - | - | 3 |
| DMTF1 | Protein coding | 13,338,261-13,377,136 | - | - | - | 10 | - | - | 2 | 1 | 7 |
| TMEM243 | Protein coding | 13,378,676-13,381,317 | - | - | - | 4 | - | - | 1 | 2 | 1 |
| CROT | Protein coding | 13,474,854-13,534,793 | - | - | - | 51 | 1 | 4 | - | 6 | 44 |
| ABCB4 | Protein coding | 13,542,668-13,614,146 | - | - | - | 8 | - | - | 1 | 1 | 6 |
| MDR1 | Protein coding | 13,644,042-13,743,053 | - | - | - | 31 | - | - | - | - | 31 |
| RUNDC3B | Protein coding | 13,803,600-13,930,733 | - | - | - | 2 | - | - | - | - | 2 |
| SLC25A40 | Protein coding | 13,930,683-13,984,374 | - | - | - | 2 | - | - | - | - | 2 |
| DBF4 | Protein coding | 13,983,320-14,019,701 | - | - | - | 5 | - | - | 1 | 2 | 2 |
| ENSCAFG00000038610 | lincRNA | 13,984,440-13,996,438 | - | - | - | 3 | - | - | 1 | 1 | 1 |
| ADAM22 | Protein coding | 14,032,475-14,250,288 | - | - | - | 2 | - | - | - | - | 2 |
| STEAP4 | Protein coding | 14,319,840-14,345,870 | - | - | - | 5 | - | - | 5 | - | - |
| ZNF804B | Protein coding | 15,146,202-15,247,064 | - | - | - | 1 | - | - | - | - | 1 |
| ENSCAFG00000037438 | lincRNA | 15,322,541-15,348,896 | - | - | - | 16 | - | - | 1 | 2 | 13 |
| ENSCAFG00000021475 | snRNA | 15,626,361-15,626,464 | - | - | - | 4 | - | - | 3 | 1 | - |
| ENSCAFG00000040353 | lincRNA | 15,711,267-15,714,286 | - | - | - | 2 | - | - | - | 2 | - |
| ENSCAFG00000034336 | lincRNA | 15,948,486-15,996,712 | - | - | - | 1 | - | - | - | - | 1 |
| STEAP1 | Protein coding | 16,074,785-16,088,182 | - | - | - | 2 | - | - | 2 | - | - |
| ENSCAFG00000036574 | lincRNA | 8,141,376-8,144,851 | - | - | - | 1 | - | - | 1 | - | - |
| ENSCAFG00000032858 | lincRNA | 8,148,588-8,152,852 | - | - | - | 1 | - | - | - | - | 1 |
| ENSCAFG00000020499 | miRNA | 8,165,107-8,165,179 | - | - | - | 1 | - | - | - | 1 | - |
| SND1 | Protein coding | 8,260,919-8,685,952 | - | - | - | 4 | - | 1 | 2 | - | 1 |
| PAX4 | Protein coding | 8,714,949-8,719,371 | - | - | - | 2 | - | - | 1 | 1 | - |
| FSCN3 | Protein coding | 8,724,865-8,730,729 | - | - | - | 1 | - | - | - | 1 | - |
| ARF5 | Protein coding | 8,728,365-8,735,331 | - | - | - | 1 | - | - | 1 | - | - |
| GCC1 | Protein coding | 8,738,782-8,742,022 | - | - | - | 1 | - | - | 1 | - | - |
| ZNF800 | Protein coding | 8,866,900-8,884,790 | - | - | - | 1 | - | - | - | - | 1 |
| GRM8 | Protein coding | 8,993,463-9,719,899 | - | - | - | 61 | - | - | - | - | 61 |

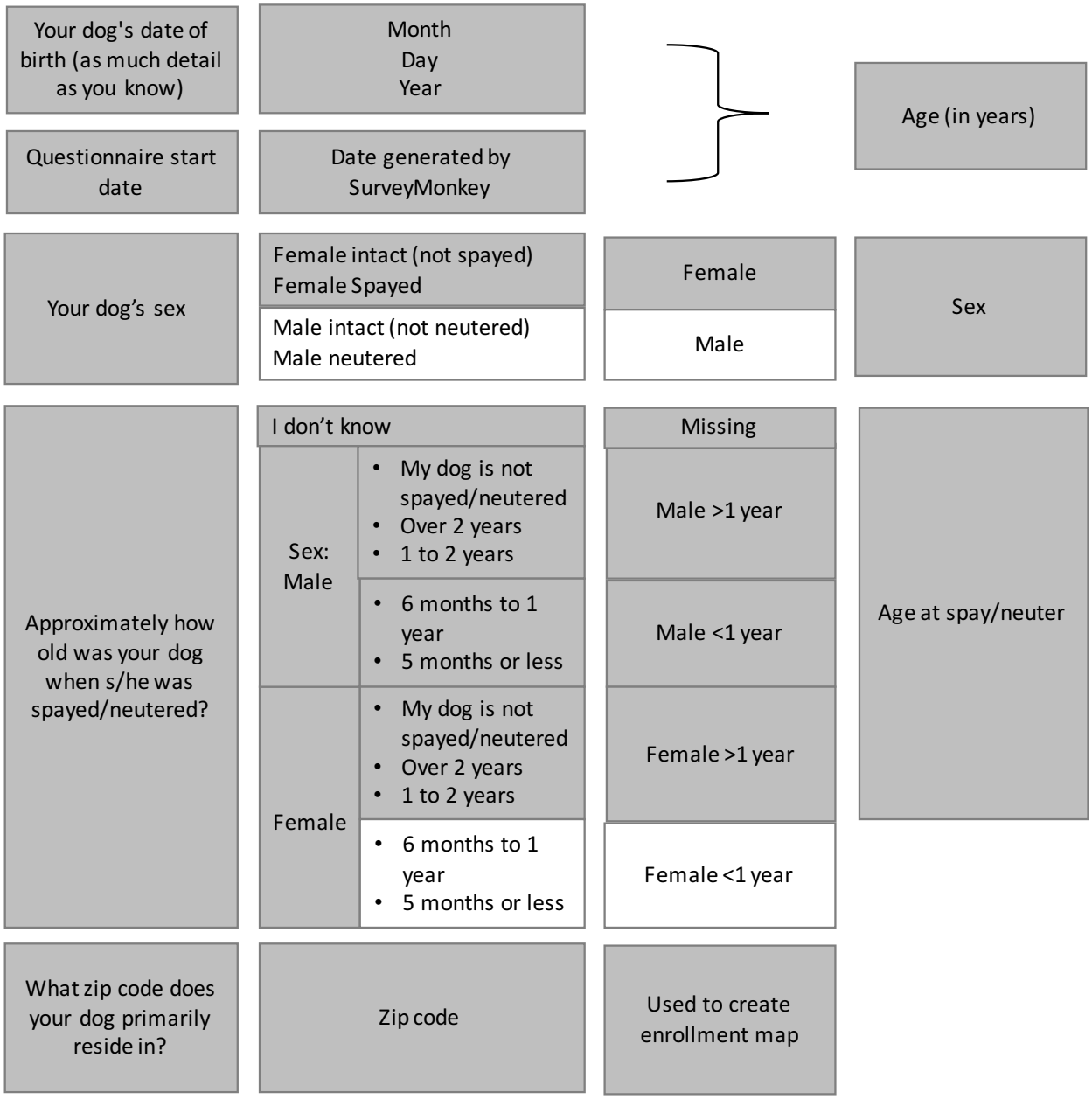
**Suppl. Table 3.5. Summary of potential regulatory elements.**

| Enhancer name | Human location | | Gene targets | Variants in location |
|---|---|---|---|---|
| GH14F080149 | chr14 | 80149793-80152533 | DIO2, LOC105370591, LOC105370593 | cfa8:52637369C>T, cfa8:52637577C>T, cfa8:52638277ATTC>A, cfa8:52638688T>G, cfa8:52638807T>C, cfa8:52638810C>T, cfa8:52638841T>C, cfa8:52638957G>A, cfa8:52638968T>C, cfa8:52639257T>C, cfa8:52639425T>C |
| GH14F080490 | chr14 | 80490741-80492445 | DIO2, PIR42122, HMGN2P2 | cfa8:52908241T>G, cfa8:52909750C>A, cfa8:52909963C>A, cfa8:52910636A>G, cfa8:52910641G>A, cfa8:52910725G>T, cfa8:52910779AGC>A, cfa8:52910782ACTTCTC>A, cfa8:52911069GAATGCTTCA>G, cfa8:52911079A>AGGGGG |
| GH14F081218 | chr14 | 81218244-81223058 | GTF2A1, CEP128, LOC101928504, SNORA79, GC14M081203 | cfa8:53562158G>A, cfa8:53564825A>C, cfa8:53569483CTT>CT, cfa8:53569932TAAC>T |
| GH14F081422 | chr14 | 81422801-81428583 | STON2, LOC100506700, DYNLL1P2 | cfa8:53721593G>T, cfa8:53724541C>A |
| GH14F081434 | chr14 | 81434277-81437366 | GTF2A1, STON2, LOC100506700, DYNLL1P2 | cfa8:53733232T>C, cfa8:53733863C>A, cfa8:53736400T>TA |
| GH14F081473 | chr14 | 81473257-81475664 | GTF2A1, SEL1L, SNORA79, CEP128, LOC100506700 | cfa8:53768127C>T, cfa8:53768170ACAAT>A, cfa8:53768224T>C, cfa8:53768739C>T, cfa8:53768749GA>G, cfa8:53769789TA>T |
| GH14F081516 | chr14 | 81516304-81520717 | SEL1L, LOC100506700 | cfa8:53808023G>C, cfa8:53812372G>A, cfa8:53812505T>G, cfa8:53812691T>C, cfa8:53812902G>C, cfa8:53812949A>G, cfa8:53812956T>A, cfa8:53812962C>A, cfa8:53813064G>C, cfa8:53813354T>C |
| GH14F081529 | chr14 | 81529402-81530399 | SEL1L, LOC100506700 | cfa8:53820166G>A, cfa8:53820327T>C, cfa8:53820384G>A |
| GH14F081532 | chr14 | 81532659-81536926 | GTF2A1, SEL1L, LINC01467 | cfa8:53822344T>C, cfa8:53825141G>C, cfa8:53825250G>A, cfa8:53825410G>GT, cfa8:53826199C>T, cfa8:53826369T>TACCC, cfa8:53826727C>T, cfa8:53826790G>A, cfa8:53826797AT>A, cfa8:53826992T>C, cfa8:53827032G>A, cfa8:53827044C>T, cfa8:53827076A>G, cfa8:53827088C>A, cfa8:53827096A>C |

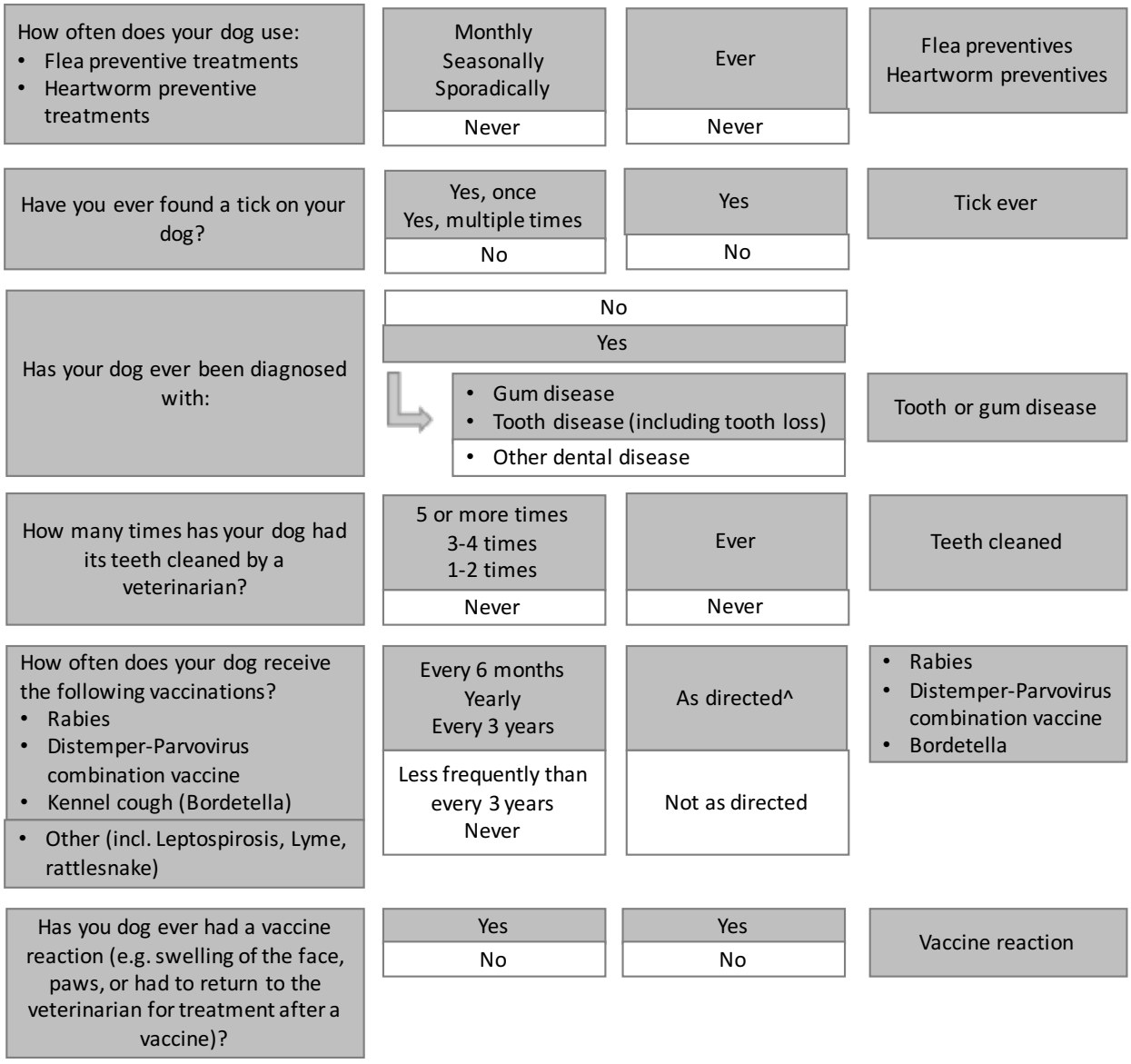
APPENDIX 3: SUPPLEMENTARY FIGURES AND TABLES FOR AIM 3

**Suppl. Figure 4.1. Collapsing of variables from questionnaire for use in analysis.**

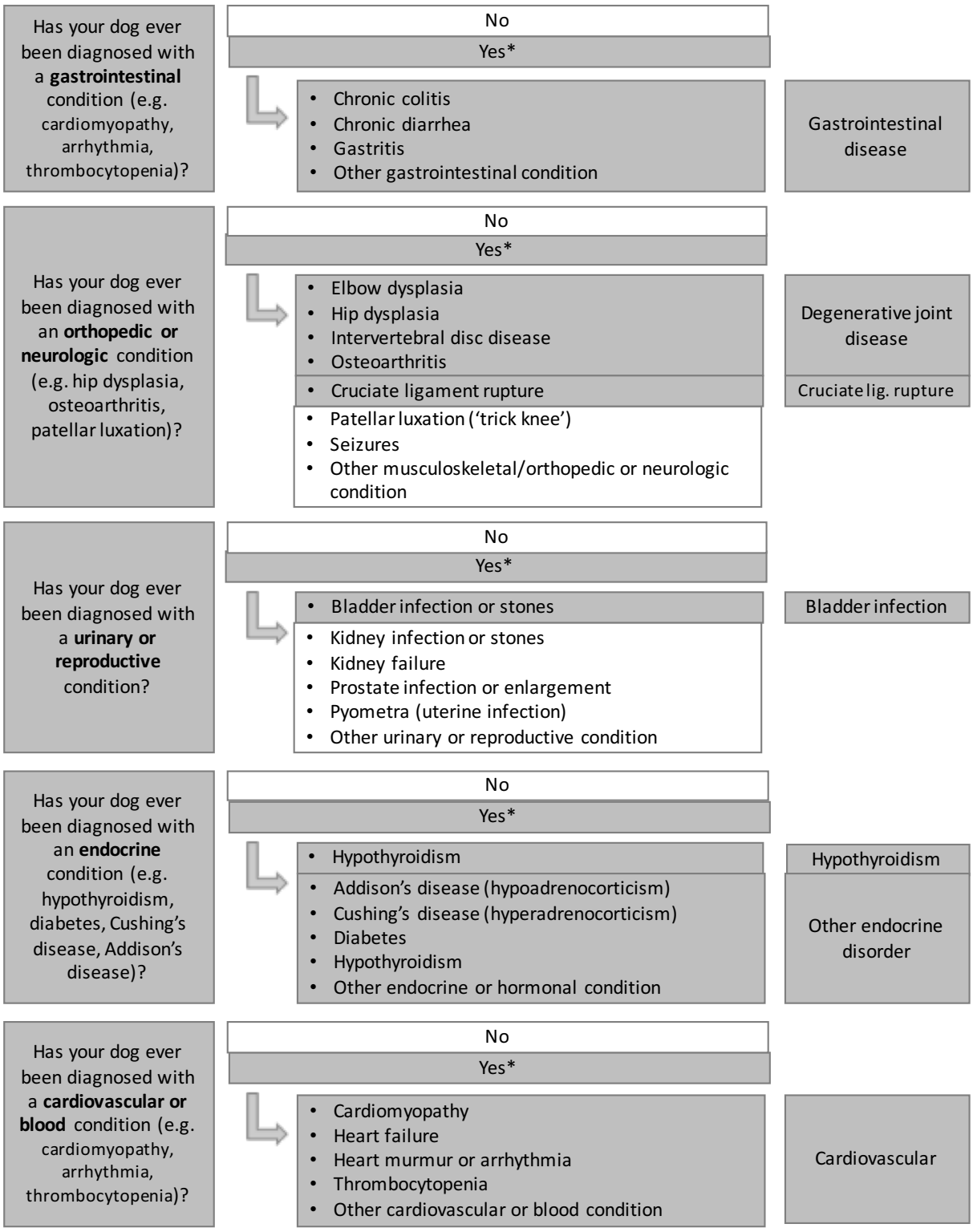
A. Signalment

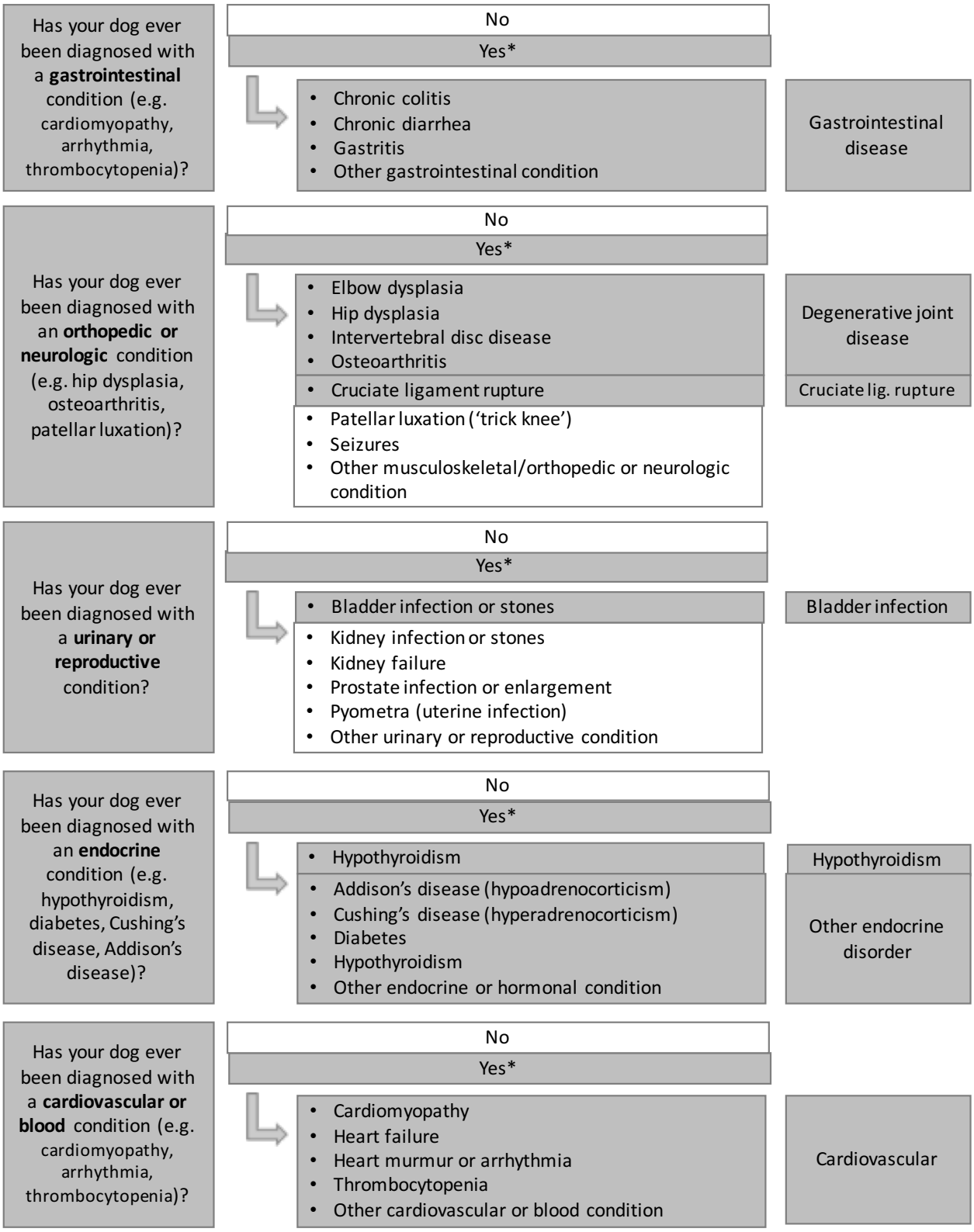


## B. Preventive care



## D. Health history




Has your dog ever been diagnosed with a **gastrointestinal** condition (e.g. cardiomyopathy, arrhythmia, thrombocytopenia)?
No
Yes*
• Chronic colitis
• Chronic diarrhea
• Gastritis
• Other gastrointestinal condition
Gastrointestinal disease
Has your dog ever been diagnosed with an **orthopedic or neurologic** condition (e.g. hip dysplasia, osteoarthritis, patellar luxation)?
No
Yes*
• Elbow dysplasia
• Hip dysplasia
• Intervertebral disc disease
• Osteoarthritis
Degenerative joint disease
• Cruciate ligament rupture
Cruciate lig. rupture
• Patellar luxation ('trick knee')
• Seizures
• Other musculoskeletal/orthopedic or neurologic condition
Has your dog ever been diagnosed with a **urinary or reproductive** condition?
No
Yes*
• Bladder infection or stones
Bladder infection
• Kidney infection or stones
• Kidney failure
• Prostate infection or enlargement
• Pyometra (uterine infection)
• Other urinary or reproductive condition
Has your dog ever been diagnosed with an **endocrine** condition (e.g. hypothyroidism, diabetes, Cushing's disease, Addison's disease)?
No
Yes*
• Hypothyroidism
Hypothyroidism
• Addison's disease (hypoadrenocorticism)
• Cushing's disease (hyperadrenocorticism)
• Diabetes
• Hypothyroidism
• Other endocrine or hormonal condition
Other endocrine disorder
Has your dog ever been diagnosed with a **cardiovascular or blood** condition (e.g. cardiomyopathy, arrhythmia, thrombocytopenia)?
No
Yes*
• Cardiomyopathy
• Heart failure
• Heart murmur or arrhythmia
• Thrombocytopenia
• Other cardiovascular or blood condition
Cardiovascular

## D. Medications, environmental exposures, and diet

| | | | |
|---|---|---|---|
| Other than for treatment of cancer, has your dog ever taken any of the following medications?<br>• Antihistamines (e.g. Diphenhydramine/Benadryl, Chlorpheniramine, hydroxyzine/atarax)<br>• Anti-inflammatories (e.g. Aspirin, Rimadyl, Previcox)<br>• Steroids (e.g. Prednisone, Dexamethasone)<br>• Antibiotics<br>• Oral immunosuppressants (e.g. azathioprine, cyclosporine) | One course<br>Multiple courses<br>Continuous use<br>Never | Ever<br>Never | Antihistamines<br>Anti-inflammatories<br>Steroids<br>Antibiotics<br>Oral immunosuppressants |
| • Non-prescription supplements or alternative therapies<br>• Other medication, including topical | Text box indicating supplement/other medication type | Ever<br>Never | Omega-3 supplements<br>Other supplements |
| On average, how much time does your dog spend on **your lawn** at home? | Over 18 hours per day<br>13-18 hours per day<br>6 -12 hours per day<br>1-5 hours per day<br><1 hour per day<br>Never | Frequent<br>Infrequent | Lawn exposure |
| On average, how much time does your dog spend in the following areas?<br>• Rural environments?<br>• Parks or other public grassy spaces? | Daily<br>Weekly<br>Monthly<br>Occasionally<br>Never | Frequent<br>Infrequent | Rural environment<br>Parks |
| On average, how often is your dog exposed to:<br>• Lawn chemicals<br>• Cigarette smoke<br>• Paints, solvents, cleaning fluids, etc.? | Daily<br>Weekly<br>Monthly<br>Occasionally<br>Never | Frequent<br>Infrequent | Lawn chemicals<br>Cigarette smoke<br>Other chemicals |
| On average, how often does your dog swim in the following areas?<br>• Ocean<br>• Irrigation ditches, ponds, or canals<br>• Lakes or streams | Daily<br>Weekly<br>Monthly<br>Occasionally<br>Never | Frequent<br>Infrequent | Ocean<br>Irrigation water<br>Lakes/streams |
| What kind of food comprises the majority of your dog's diet? | Commercial dry food (kibble)<br>Commercial canned food<br>Commercial "fresh" or frozen food<br>Home prepared food – cooked<br>Raw diet | | |

**Suppl. Table 4.1. Modeling results for case vs control sensitivity analyses.** *Final model* indicates the model reported as our main results (n=140 cases, n=147 controls). *Age sensitivity analysis* removed 20 cases that were <9 years of age (n=120 cases, n=147 controls). *Cancer sensitivity analysis* removed both cases and controls with a history of or concurrent non-lymphoma cancer (n=119 cases, n=111 controls). % change indicates the percent change in the OR estimate compared to the final model.

| | | Final model | | | | Age sensitivity analysis | | | | | Cancer sensitivity analysis | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | OR | (95% CI) | Indiv p-value | Type3 p-value | OR | (95% CI) | Indiv p-value | Type3 p-value | % change | OR | (95% CI) | Indiv p-value | Type3 p-value | % change |
| Age at | Female >1yr | 0.34 | (0.16 - 0.74) | 0.007 | | 0.39 | (0.17 - 0.85) | 0.018 | | 12% | 0.39 | (0.17 - 0.90) | 0.028 | | 12% |
| spay/neuter | Male <1yr | 0.62 | (0.3 - 1.3) | 0.206 | | 0.65 | (0.30 - 1.40) | 0.274 | | 5% | 0.74 | (0.32 - 1.70) | 0.474 | | 19% |
| (vs Female <1yr) | Male >1yr | 0.93 | (0.4 - 2.16) | 0.862 | | 0.91 | (0.37 - 2.22) | 0.827 | | 2% | 0.84 | (0.33 - 2.11) | 0.704 | | 10% |
| | Missing | 1.44 | (0.46 - 4.46) | 0.530 | 0.031 | 1.72 | (0.55 - 5.42) | 0.352 | 0.054 | 20% | 1.26 | (0.36 - 4.43) | 0.723 | 0.175 | 13% |
| DHPP | As directed (vs not) | 3.21 | (1.58 - 6.51) | 0.001 | 0.001 | 3.03 | (1.47 - 6.24) | 0.003 | 0.003 | 6% | 3.45 | (1.58 - 7.55) | 0.002 | 0.002 | 8% |
| Bladder infection | Yes (vs no) | 3.49 | 0.96 - 12.67) | 0.057 | 0.057 | 2.95 | 0.77 - 11.35) | 0.116 | 0.116 | 16% | 2.94 | 0.64 - 13.36) | 0.164 | 0.164 | 16% |
| Mange | Yes (vs no) | 5.50 | (1.44 - 21.1) | 0.013 | 0.013 | 5.78 | 1.49 - 22.39) | 0.011 | 0.011 | 5% | 11.52 | 1.36 - 97.29) | 0.025 | 0.025 | 109% |
| Eye disease | Yes (vs no) | 2.25 | (0.97 - 5.22) | 0.059 | 0.059 | 2.49 | (1.06 - 5.88) | 0.037 | 0.037 | 11% | 1.52 | (0.58 - 3.94) | 0.394 | 0.394 | 33% |
| Gastrointestinal disease | Yes (vs no) | 2.38 | (0.98 - 5.76) | 0.055 | 0.055 | 2.20 | (0.90 - 5.40) | 0.085 | 0.085 | 7% | 3.30 | (1.13 - 9.65) | 0.030 | 0.030 | 39% |
| Hypothyroidism | Yes (vs no) | 0.25 | (0.1 - 0.66) | 0.005 | 0.005 | 0.27 | (0.10 - 0.74) | 0.011 | 0.011 | 8% | 0.40 | (0.14 - 1.15) | 0.089 | 0.089 | 60% |
| Omega-3 | Yes | 0.29 | (0.13 - 0.63) | 0.002 | 0.007 | 0.27 | (0.12 - 0.62) | 0.002 | | 6% | 0.32 | (0.14 - 0.79) | 0.012 | 0.026 | 12% |
| supplements (vs | Missing | 0.88 | (0.44 - 1.73) | 0.704 | | 0.94 | (0.47 - 1.88) | 0.857 | | 7% | 1.20 | (0.56 - 2.56) | 0.643 | | 37% |

**Suppl. Table 4.2. Modeling results for subclinical vs control sensitivity analysis.** *Final model* indicates the model reported as our main results (n=221 subclinicals, n=147 controls). *Cancer sensitivity analysis* removed both subclinicals and controls with a history of or concurrent non-lymphoma cancer (n=166 subclinicals, n=111 controls). % change indicates the percent change in the OR estimate compared to the final model.

| | | Final model | | | | Cancer sensitivity analysis | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | OR | (95% CI) | Indiv p-value | Type3 p-value | OR | (95% CI) | Indiv p-value | Type3 p-value | % change |
| Age at spay/neuter | Female >1yr | 1.07 | (0.59 - 1.95) | 0.831 | 0.050 | 1.42 | (0.71 - 2.87) | 0.325 | | 33% |
| (vs Female <1yr) | Male <1yr | 0.82 | (0.42 - 1.6) | 0.566 | | 0.91 | (0.40 - 2.05) | 0.815 | | 10% |
| | Male >1yr | 2.22 | (1.13 - 4.37) | 0.021 | | 2.32 | (1.07 - 5.03) | 0.034 | | 4% |
| | Missing | 1.16 | (0.37 - 3.68) | 0.800 | | 1.39 | (0.38 - 5.05) | 0.615 | 0.142 | 20% |
| Bladder infection | Yes (vs no) | 5.10 | (1.9 - 13.69) | 0.001 | 0.001 | 5.39 | (1.52 - 19.05) | 0.009 | 0.009 | 6% |
| Eye disease | Yes (vs no) | 1.92 | (0.99 - 3.75) | 0.056 | 0.001 | 2.10 | (0.98 - 4.48) | 0.056 | 0.056 | 9% |

**Suppl. Figure 4.2. Questionnaire used in aim 3 analysis.**

## Dog Health Survey

### Section 1: General Information

**1. Your name**

First

Last

Email

**2. Your dog's name**

**3. Your dog's date of birth (as much detail as you know)**

Month

Day

Year

**4. Your dog's sex**

- Female intact (not spayed)
- Female spayed
- Male intact (not neutered)
- Male neutered

**5. Approximately how old was your dog when s/he was spayed/neutered?**

- 5 months or less
- 6 months to 1 year
- 1 to 2 years
- Over 2 years
- I don't know
- My dog is not spayed/neutered

**6. If your dog is a spayed female, was she spayed before or after her first estrus ("heat") cycle?**

- ○ Before
- ○ After
- ○ I don't know
- ○ My dog is not a spayed female

**7. When did you acquire your dog?**

Month

Year

**8. Where did you acquire your dog?**

City

State

Zip

**9. What zip code does your dog primarily reside in?**

## Section 2: Health Information - Preventive Care

**10. How often do you use flea and/or heartworm preventative treatments?**

| | Monthly | Seasonally | Sporadically | Never |
|---|---|---|---|---|
| Flea | ○ | ○ | ○ | ○ |
| Heartworm | ○ | ○ | ○ | ○ |

**11. Have you ever found a tick on your dog?**

- ○ Yes, once
- ○ Yes, multiple times
- ○ No

**12. Has your dog ever been diagnosed with any of the following? Check all that apply**

- ☐ Gum disease
- ☐ Tooth disease (including tooth loss)
- ☐ Other dental disease
- ☐ None of the above

Please provide detail about type of dental disease

**13. How many times has your dog had its teeth cleaned by a veterinarian?**

- ○ 1-2 times
- ○ 3-4 times
- ○ 5 or more times
- ○ Never

**14. How often does your dog receive the following vaccinations?**

**Note: rabies and distemper are considered core vaccines and commonly given at 1 or 3 year intervals**

| | Every 6 months | Yearly | Every 3 years | Less frequently than every 3 years | Never |
|---|---|---|---|---|---|
| Rabies | ○ | ○ | ○ | ○ | ○ |
| Distemper-Parvovirus Combination Vaccine (DHPP or DHLPPC) | ○ | ○ | ○ | ○ | ○ |
| Kennel Cough (Bordetella) | ○ | ○ | ○ | ○ | ○ |
| Leptospirsosis | ○ | ○ | ○ | ○ | ○ |
| Lyme | ○ | ○ | ○ | ○ | ○ |
| Rattlesnake | ○ | ○ | ○ | ○ | ○ |
| Other | ○ | ○ | ○ | ○ | ○ |

If other, please specify

**15. Has your dog ever had a vaccine reaction (e.g. swelling of the face, paws, or had to return to the veterinarian for treatment after a vaccine)?**

○ No

○ Yes (please specify vaccine type and approximate age of dog)

## Section 2: Health Information - Health History

**16. Has your dog ever been diagnosed with CANCER?**

- ○ Yes
- ○ No

## Section 2: Health Information - Cancer

**17. Please indicate any CANCER your dog has been diagnosed with and the age at diagnosis.**

| | <1 year | 1-3 years | 4-6 years | 7-10 years | Over 10 years | Diagnosed with this condition, but can't remember age |
|---|---|---|---|---|---|---|
| Breast or mammary cancer | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Hemangiosarcoma | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Histiocytoma | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Lymphoma/Leukemia | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Mast cell tumor | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Melanoma | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Osteosarcoma | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Squamous cell carcinoma | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Transitional cell carcinoma | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Other cancer | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

If other, please specify type and age at diagnosis

**18. Is your dog CURRENTLY receiving any of the following treatments for cancer? Check all that apply**

- ☐ Chemotherapy
- ☐ Radiation
- ☐ Steroids (e.g. Prednisone)
- ☐ Other therapy

Please elaborate on treatment and cancer type

**19. Has your dog PREVIOUSLY received any of the following treatments for cancer? Check all that apply**

- [ ] Chemotherapy
- [ ] Radiation
- [ ] Steroids (e.g. Prednisone)
- [ ] Other therapy

Please elaborate on treatment and cancer type

## Section 2: Health Information - Health History

**20. Has your dog ever been diagnosed with an EYE, EAR, NOSE, or THROAT condition (e.g. ear infection, uveitis, cataracts)?**

- ○ Yes
- ○ No

## Section 2: Health Information - ENT

**21. Please indicate any EAR, NOSE, THROAT, or EYE conditions your dog has been diagnosed with and the age at diagnosis.**

| | <1 year | 1-3 years | 4-6 years | 7-10 years | Over 10 years | Diagnosed with this condition, but can't remember age |
|---|---|---|---|---|---|---|
| Cataracts | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Ear infection (Otitis externa) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Glaucoma | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Uveitis | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Progressive retinal atrophy or degeneration | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Sudden onset blindness | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Other ear, nose, throat, or eye condition | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

If other, please specify type and age at diagnosis

## Section 2: Health Information - Health History

**22. Has your dog ever been diagnosed with an INFECTIOUS DISEASE (e.g. worms, parvovirus, ehrlichiosis, kennel cough)?**

- ○ Yes
- ○ No

## Section 2: Health Information - Infectious Disease

**23. Please indicate any INFECTIOUS DISEASES your dog has been diagnosed with and the age at diagnosis.**

| | <1 year | 1-3 years | 4-6 years | 7-10 years | Over 10 years | Diagnosed with this condition, but can't remember age |
|---|---|---|---|---|---|---|
| Giardia | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Intestinal disease suspected to be infectious | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Kennel Cough (Tracheobronchitis) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Leptospirosis | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Parvovirus | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Tick-born disease (e.g. Ehrlichiosis, Lyme, Rocky Mountain Spotted Fever) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Worms (e.g. heartworm, tapeworm) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Other infectious disease | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

If other, please specify type and age at diagnosis

## Section 2: Health Information - Health History

**24. Has your dog ever been diagnosed with a SKIN condition (e.g. itchy skin, allergies, skin infection, mange)?**

- ○ Yes
- ○ No

## Section 2: Health Information - Skin

**25. Please indicate any SKIN conditions your dog has been diagnosed with and the age at diagnosis.**

| | <1 year | 1-3 years | 4-6 years | 7-10 years | Over 10 years | Diagnosed with this condition, but can't remember age |
|---|---|---|---|---|---|---|
| 'Hot Spots' | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Itchy skin/allergies | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Mange, scabies, or mites | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Pyoderma (skin infection) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Other skin condition | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

If other, please specify type and age at diagnosis

## Section 2: Health Information - Health History

**26. Has your dog ever been diagnosed with a GASTROINTESTINAL condition (e.g. chronic diarrhea, colitis)?**

- ○ Yes
- ○ No

## Section 2: Health Information - GI

**27. Please indicate any GASTROINTESTINAL conditions your dog has been diagnosed with and the age at diagnosis.**

| | <1 year | 1-3 years | 4-6 years | 7-10 years | Over 10 years | Diagnosed with this condition, but can't remember age |
|---|---|---|---|---|---|---|
| Chronic colitis | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Chronic diarrhea | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Gastritis | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Other gastrointestinal condition | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

If other, please specify type and age at diagnosis

## Section 2: Health Information - Health History

**28. Has your dog ever been diagnosed with an ORTHOPEDIC or NEUROLOGIC condition (e.g. hip dysplasia, osteoarthritis, patellar luxation)?**

- ○ Yes
- ○ No

## Section 2: Health Information - Orthopedic

**29. Please indicate any ORTHOPEDIC or NEUROLOGIC conditions your dog has been diagnosed with and the age at diagnosis.**

| | <1 year | 1-3 years | 4-6 years | 7-10 years | Over 10 years | Diagnosed with this condition, but can't remember age |
|---|---|---|---|---|---|---|
| Cruciate ligament rupture | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Elbow dysplasia | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Hip dysplasia | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Invertebral disc disease | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Osteoarthritis | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Patellar luxation ('trick knee') | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Seizures | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Other musculoskeletal/orthopedic or neurologic condition | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

If other, please specify type and age at diagnosis

## Section 2: Health Information - Health History

**30. Has your dog ever been diagnosed with a URINARY or REPRODUCTIVE condition?**

- ○ Yes
- ○ No

## Section 2: Health Information - Urinary/Repro

**31. Please indicate any URINARY OR REPRODUCTIVE conditions your dog has been diagnosed with and the age at diagnosis.**

| | <1 year | 1-3 years | 4-6 years | 7-10 years | Over 10 years | Diagnosed with this condition, but can't remember age |
|---|---|---|---|---|---|---|
| Bladder infection or stones | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Kidney infection or stones | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Kidney failure | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Prostate infection or enlargement | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Pyometra (uterine infection) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Other urinary or reproductive condition | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

If other, please specify type and age at diagnosis

## Section 2: Health Information - Health History

**32. Has your dog ever been diagnosed with an ENDOCRINE condition (e.g. hypothyroidism, diabetes, Cushing's disease, Addison's disease)?**

- ○ Yes
- ○ No

## Section 2: Health Information - Endocrine

**33. Please indicate any ENDOCRINE conditions your dog has been diagnosed with and the age at diagnosis.**

| | <1 year | 1-3 years | 4-6 years | 7-10 years | Over 10 years | Diagnosed with this condition, but can't remember age |
|---|---|---|---|---|---|---|
| Addison's disease (hypoadrenocorticism) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Cushing's disease (hyperadrenocorticism) | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Diabetes | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Hypothyroidism | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Other endocrine or hormonal condition | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

If other, please specify type and age at diagnosis

## Section 2: Health Information - Health History

**34. Has your dog ever been diagnosed with a CARDIOVASCULAR or BLOOD condition (e.g. cardiomyopathy, arrhythmia, thrombocytopenia)?**

- ○ Yes
- ○ No

## Section 2: Health Information - Cardio

**35. Please indicate any CARDIOVASCULAR or BLOOD conditions your dog has been diagnosed with and the age at diagnosis.**

| | <1 year | 1-3 years | 4-6 years | 7-10 years | Over 10 years | Diagnosed with this condition, but can't remember age |
|---|---|---|---|---|---|---|
| Cardiomyopathy | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Heart failure | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Heart murmur or arrhythmia | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Thrombocytopenia | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |
| Other cardiovascular or blood condition | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

If other, please specify type and age at diagnosis

## Section 2: Health Information - Health History

**36. Additional significant medical history**

**37. Other than for treatment of cancer, has your dog ever taken any of the following medications?**

| | Never | One course | Multiple courses | Continuous use |
|---|---|---|---|---|
| ANTIHISTAMINES (e.g. Diphenhydramine/Benadryl, Chlorpheniramine, hydroxyzine/atarax) | ○ | ○ | ○ | ○ |
| ANTI-INFLAMMATORIES (e.g. Aspirin, Rimadyl, Previcox) | ○ | ○ | ○ | ○ |
| STEROIDS (e.g. Prednisone, Dexamethasone) | ○ | ○ | ○ | ○ |
| ANTIBIOTICS | ○ | ○ | ○ | ○ |
| ORAL IMMUNOSUPPRESSANTS (e.g. azathioprine, cyclosporine) | ○ | ○ | ○ | ○ |
| NON-PRESCRIPTION SUPPLEMENTS or alternative therapies | ○ | ○ | ○ | ○ |
| OTHER medication, including topical | ○ | ○ | ○ | ○ |

Please specify other medications or supplements (if applicable)

## Section 3: Environmental Information

**38. On average, how much time does your dog spend on your lawn at home?**

- ○ I do not have a lawn at home
- ○ Less than 1 hour per day
- ○ 1 - 5 hours per day
- ○ 6 - 12 hours per day
- ○ 13 - 18 hours per day
- ○ Over 18 hours per day

**39. On average, how much time does your dog spend in the following areas?**

| | Daily | Weekly | Monthly | Occasionally | Never | Other |
|---|---|---|---|---|---|---|
| Rural environments | ○ | ○ | ○ | ○ | ○ | ○ |
| Parks or other public grassy spaces | ○ | ○ | ○ | ○ | ○ | ○ |

If other, please specify

**40. On average, how often is your dog exposed to**

| | Daily | Weekly | Monthly | Occasionally | Never | Other |
|---|---|---|---|---|---|---|
| Lawn chemicals | ○ | ○ | ○ | ○ | ○ | ○ |
| Cigarette smoke | ○ | ○ | ○ | ○ | ○ | ○ |
| Paints, solvents, cleaning fluids, etc | ○ | ○ | ○ | ○ | ○ | ○ |

If other, please specify

**41. On average, how often does your dog swim in the following areas?**

| | Daily | Weekly | Monthly | Occasionally | Never | Other |
|---|---|---|---|---|---|---|
| Ocean | ○ | ○ | ○ | ○ | ○ | ○ |
| Irrigation ditches, ponds, or canals | ○ | ○ | ○ | ○ | ○ | ○ |
| Lakes or streams | ○ | ○ | ○ | ○ | ○ | ○ |

If other, please specify

**42. What kind of food comprises the majority of your dog's diet? Check all that apply**

- ☐ Commercial dry food (kibble)
- ☐ Commercial canned food
- ☐ Commercial "fresh" or frozen food
- ☐ Home prepared food - cooked
- ☐ Raw diet

If commercial, please specify brand and variety

## Feedback

**43. Please provide feedback about the questionnaire (optional)**

**Thank you very much for your time!**