DISSERTATION

INSIGHTS INTO THE CELL OF ORIGIN, PATHOGENESIS, AND TRANSLATIONAL POTENTIAL OF CANINE PERIPHERAL T-CELL LYMPHOMA

Submitted by

Eileen Owens

Department of Microbiology, Immunology & Pathology

In partial fulfilment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2025

Doctoral Committee:

Advisor: Anne Avery

Randall Basaraba
Dawn Duval
Amy MacNeill
Daniel Regan

ABSTRACT

INSIGHTS INTO THE CELL OF ORIGIN, PATHOGENIESIS, AND TRANSLATIONAL POTENTIAL OF CANINE PERIPHERAL T-CELL LYMPHOMA

Peripheral T-cell lymphoma (PTCL) refers to a heterogenous group of T-cell neoplasms in humans and dogs with short survival times and poor treatment responses. The clinical features and immunophenotype of PTCL in dogs closely resemble PTCL, not otherwise specified (PTCL-NOS), the most common and most poorly understood subtype of PTCL in humans. This has led to interest in their potential as a preclinical model for this disease. Because these are naturally occurring tumors in dogs, they offer some unique advantages as a preclinical model compared to traditional xenograft and genetically engineered mouse models, including the ability to study these cancers in an immunocompetent host with a more comparable tumor microenvironment, the potential to investigate the genetic origins of these tumors in hosts with similar genetic diversity to humans, and the consideration of environmental risk factors, since humans and dogs largely share the same environments. One remaining barrier to this proposed canine model is our limited understanding of how the molecular features of canine PTCL compare to human PTCL-NOS. In humans, gene expression and mutational profiling studies have unveiled prognostically significant molecular subtypes of PTCL-NOS, identified enriched oncogenic signaling pathways, recognized mutations in key tumor suppressor genes and oncogenes, and discovered large structural variants such as fusion genes that contribute to the pathogenesis of PTCL-NOS and represent opportunities for novel targeted therapies. However, similar gene expression and mutational profiling studies in canine PTCL are lacking.

We hypothesized that canine PTCL would exhibit a similar global gene expression profile to PTCL-NOS, would be enriched for the same oncogenic signaling pathways, and would share some of the most common gene mutations of human PTCL-NOS. We tested this hypothesis by performing bulk RNA-sequencing, differential gene expression analysis, and variant analysis on canine PTCL, normal canine nodal lymphocytes, and normal canine thymocytes. We discovered that the gene expression profile of the most common $CD3^+CD5^{+/-}CD4^+CD8^-$ immunophenotype of canine PTCL closely resembles human PTCL-NOS—particularly the *GATA3*-expressing subtype associated with an inferior prognosis—and canine PTCL often exhibits mutations in the same genes commonly mutated in human PTCL-NOS, including *DNMT3A, TP53, TET2,* and *PTEN*.

The cell of origin of human PTCL-NOS remains a subject of ongoing speculation. The observation that subsets of human PTCL-NOS upregulate either *TBX21* or *GATA3*—transcription factors controlling the differentiation of naïve T cells to mature T helper 1 (Th1) or T helper 2 (Th2) cells, respectively—led to suggestions that these tumors are derived from their respective normal mature T-helper cell counterparts. However, more recent studies have demonstrated that *GATA3* is an oncogene upregulated by a variety of neoplasms, including immature precursor T-cell neoplasms and even non-hematopoietic neoplasms, and its expression in GATA3-PTCL is associated with promotion of general T-cell survival and proliferation rather than Th2 differentiation. Additionally, mature T cells have been shown to be highly resistant to oncogenic transformation, suggesting that even seemingly mature T-cell neoplasms may actually arise from earlier precursors. Given the similarities between human PTCL-NOS and canine $CD4^+$ PTCL, investigations into the cell of origin in dogs may offer additional insight into the cell of origin in human PTCL-NOS. One previous study identified decreased surface expression of CD25 and MHC class II in canine PTCL, molecules normally associated with T-cell maturity and activation, suggesting that canine PTCL may arise from an earlier T-cell precursor. Recently, an RNA-sequencing study comparing the gene expression

profile of canine T-zone lymphoma, a more indolent form of T-cell lymphoma in dogs, to gene signatures of human and murine naïve and activated T-cell subsets successfully identified the cell of origin for this disease as a mature, previously activated αβ T cell. Therefore, we sought to employ similar methods to test our hypothesis that the cell of origin of canine $CD4^+$ PTCL is a thymic precursor cell. Additionally, to help overcome some of the limitations of using publicly available data in other species, we sought to develop a single-cell transcriptomic atlas of normal canine thymic and lymph node tissue to identify key changing genes throughout T-cell development and differentiation in dogs and evaluate those expression signatures in canine $CD4^+$ PTCL. We found that canine $CD4^+$ PTCL upregulated several markers of immaturity, including *CD34, KIT, DNTT*, and *CCR9*. Canine $CD4^+$ PTCL was enriched for gene signatures associated with immature murine and human thymocytes compared to more mature T-cell subsets. Finally, based on data derived from our single-cell atlas of normal thymus and lymph node, canine $CD4^+$ PTCL tended to upregulate genes expressed early in canine thymocytes and downregulate genes expressed as canine T cells progressed to more mature stages of development. Taken together, these findings support an immature thymic precursor cell of origin for canine $CD4^+$ PTCL.

For a subset of human PTCL-NOS, a type of mutation known as *fusion genes* play a role in tumorigenesis and offer an opportunity for targeted therapies. These mutations are characterized by the juxtaposition of two previously independent genes, which can then be transcribed and potentially translated together to oncogenic effect. In small subsets of human PTCL-NOS, the presence of a fusion gene drives malignant transformation through upregulation of specific oncogenic signaling pathways, and inhibitors of these pathways have shown promise as an alternative treatment option in these tumors. Despite their potential as actionable drivers in many hematologic cancers, fusion genes have not been previously investigated or described as a feature

of canine PTCL. We hypothesized that recurrent fusion genes could be identified in canine $CD4^+$ PTCL, and that some would correlate with enrichment for particular oncogenic pathways.

To test this hypothesis, we utilized fusion calling algorithms for bulk RNA-sequencing data and long-read DNA sequencing techniques to investigate expression of fusion mRNA transcripts and the presence of large chromosomal aberrations, respectively, in canine $CD4^+$ PTCL. We identified 13 recurrent fusion genes at the mRNA level, 3 of which were tumor-specific and not detected in dogs without lymphoma. 2 of these 3 PTCL-specific fusions, *TOX2-LMO4* and *PER1-EIF5A*, were associated with enrichment for TNF and NF-κB signaling and greater enrichment for elements of the tumor microenvironment, respectively, suggesting possible mechanisms by which these fusions could contribute to $CD4^+$ PTCL. Long-read sequencing additionally revealed 3 fusion genes in $CD4^+$ PTCL DNA that were also predicted in bulk RNA-seq data: *STAG2-SH2D1A, PLEKHA5-AEBP2,* and *SERPINB5-ENSCAFG00000031329*. The evidence of concurrent chromosomal aberrations and fusion transcript expression for these candidates make them stronger contenders for potential driver mutations that should be explored in future studies.

Taken together, the work in this dissertation supports that canine $CD4^+$ PTCL is a naturally occurring neoplasm in dogs that arises from a thymic precursor cell of origin and closely resembles the GATA3-expressing subtype of human PTCL-NOS. Additionally, similar to human PTCL-NOS, recurrent fusion genes are a feature of canine $CD4^+$ PTCL that may arise from a combination of structural chromosomal variants and alternative RNA splicing events, and these fusions may contribute to the pathogenesis of $CD4^+$ PTCL through activation of certain oncogenic signaling pathways. The shared features between canine $CD4^+$ PTCL and human PTCL-NOS represent potential opportunities to explore novel therapies in the dog that may be translated to the human disease.

ACKNOWLEDGEMENTS

The work presented in this dissertation was made possible by funding from the NIH-T32OD010437 Biomedical Research Training for Veterinarians award and the Colorado State University Clinical Hematopathology Laboratory.

I would first like to thank both my PhD advisor, Anne Avery, and my residency mentor, Allison Watson, for their guidance and mentorship throughout this combined training program. You both exemplify the passion, dedication, and tenacious curiosity that I aspire to both as a researcher and a pathologist. Thank you for all your support, troubleshooting, and advocacy for me throughout the last six years.

I also must acknowledge my incredible colleagues in the Clinical Hematopathology Laboratory, who trained me in many of the methodologies required for this project and assisted me with the collection of samples, execution of assays, and data analysis. A huge thank-you to Adam Harris and Evan Conaway for all the scripts you've shared throughout the last few years and the guidance you offered me as a new PhD student trying to navigate the new and intimidating world of bioinformatics. Adam, thank you especially for the time and effort you dedicated to our single-cell collaboration. Janna Yoshimoto, Rob Burnett, and Emily Rout: Thank you for all the times you trained me on a new protocol or answered all my annoying emails about where to find things in the lab. I truly would not have made it without you. Jeremy Dossey and Mike Agnew, thank you for your assistance with running CE on nearly 3000 PCR reactions, even after I'd insisted the previous plate was "the last one" (it never was). Sara Cook, it's been so rewarding to get to see where you're taking this project next, and I look forward to our continued collaboration.

To the members of my graduate committee, Randall Basaraba, Dawn Duval, Amy MacNeill, and Dan Regan: Thank you for devoting your time and support to me and this project. The interdisciplinary insight and fresh perspectives you lent to this project were invaluable.

Finally, a huge thank-you to my family, friends, and my fellow anatomic pathology residents. I am so thankful to be surrounded by people who love and support me so much.

TABLE OF CONTENTS

INTRODUCTION

Human peripheral T-cell lymphoma

Overview

Peripheral T-cell lymphoma (PTCL) refers to a poorly understood heterogeneous group of non-Hodgkin lymphomas arising from T lymphocytes. While these tumors are diverse in their clinical presentation and histologic and immunophenotypic features, they are similar in their aggressive clinical course, poor responses to treatment, and short survival times.[1]

In general, the incidence of PTCL increases with age, with the exception of some subtypes like ALK^+ anaplastic large cell lymphoma that have a higher incidence in the first three decades of life.[1,2] Males tend to be overrepresented, to ratios of up to 3:1 in some subtypes.[3] Incidence of PTCL also varies with geographical region: in Western nations, PTCL accounts for 10-15% of all non-Hodgkin lymphomas, but 15-20% of non-Hodgkin lymphomas in Asia, likely due to geographic differences in population genetics and/or virus epidemiology. For example, adult T-cell lymphoma/leukemia, a subtype of PTCL caused by human T-lymphotropic virus type 1, accounts for 25% of PTCL in Asia but only 1-2% of PTCL in Europe and North America.[3]

Clinical outcomes for PTCL are generally poor, evidenced by the <30% 5-year survival rate.[1] With the exception of ALK^+ anaplastic large cell lymphoma, >75% of PTCL patients either fail to respond to standard first-line CHOP (cyclophosphamide, doxorubicin, vincristine, and prednisone) chemotherapy, or relapse within 2 years. Even with implementation of second-line treatments like epigenetic modifiers, less than one-third of patients respond, and progression-free survival is poor.[1] Much of the effort to better characterize the subtypes and molecular features of PTCL has been driven by the goal of uncovering options for alternative, more targeted therapies that may help address these dismal treatment outcomes.

Subtypes of PTCL

Given their heterogeneity, subtyping PTCL by clinical and immunophenotypic features has proven challenging. Recent advancements in gene expression and mutational profiling have yielded over 30 subtypes of PTCL currently recognized by the World Health Organization (WHO),[4] a substantial increase from the 11 entities originally described in the 1994 Revised European American Lymphoma classification that laid the groundwork for the WHO classification system.[5,6]

Broadly, PTCL entities can be subdivided by primary anatomic site. PTCLs arising in the peripheral lymph nodes include nodal T-follicular helper cell lymphoma (nTFHL); peripheral T-cell lymphoma, not otherwise specified (PTCL-NOS); and anaplastic large cell lymphoma (ALCL). Mycosis fungoides, primary cutaneous anaplastic large cell lymphoma (PC-ALCL), and Sezary syndrome comprise the cutaneous PTCLs, while adult T-cell lymphoma/leukemia (ATLL) represents a disseminated or leukemic form. Other extranodal PTCLs include extranodal NK-/T-cell lymphoma, nasal type (ENKTCL); enteropathy associated T-cell lymphoma (EATL) and monomorphic epitheliotropic intestinal T-cell lymphoma (MEITL); and hepatosplenic T-cell lymphoma (HSTL).[4,7,8]

Beyond anatomic site, subtyping of PTCL is largely based on the presumed cell of origin. For example, the further stratification of nodal PTCLs resembles the different terminal stages of differentiation for T helper cells. nTFHL encompasses three malignancies that all appear to derive from T follicular helper (Tfh) cells: angioimmunoblastic T-cell lymphoma (AITL), nTFHL-follicular type (nTFHL-F), and nTFHL, not otherwise specified (nTFHL-NOS).[9] These tumors express multiple Tfh cell markers, overexpress Tfh-associated genes, and have a transcriptome predominated by overexpression of B-cell- and follicular dendritic cell-related genes.[1,10,11] Histologically, they exhibit a distinctive pattern of prominent vascular proliferation and a follicular dendritic cell network.[1] ALCL is characterized by proliferations of large, highly pleomorphic $CD30^+$

cells,[1] and the ALK$^+$ variety of ALCL upregulates Th17-cell-associated gene transcripts and downregulates transcripts of T-cell receptor (TCR) components.[10] Finally, the identification of the TBX21-PTCL and GATA3-PTCL molecular variants of nodal PTCL-NOS, discussed more extensively below, has led to some acceptance of a possible T helper 1 (Th1) or T helper 2 (Th2) cell of origin for these tumors, although this classification system has not been widely adopted and remains to be recognized by the WHO.[4,12]

PTCL-NOS is the most common of all PTCL subtypes, representing approximately one-third of PTCL cases diagnosed globally.[3] Incidentally, this subtype is also the most poorly understood, largely functioning as a "catch-all" diagnosis of exclusion for PTCL.[4] Recently, gene expression profiling has helped further divide the ambiguous PTCL-NOS into prognostically significant variants, TBX21-PTCL and GATA3-PTCL, with the GATA3-PTCL variant being associated with lower overall survival times.[13] The TBX21-PTCL variant is characterized by enrichment for features of T helper 1 (Th1) differentiation of helper T-cells, such as increased expression of IFN-γ and NF-κB, and enrichment for cytotoxic T-cell gene signatures.[13] The GATA3-PTCL variant, on the other hand, is characterized by strong expression of the transcription factor GATA3,[13] which is essential for T-cell development and for promoting T helper type 2 (Th2) differentiation of helper T cells;[14] enrichment for cellular proliferation pathways MYC and PI3K/AKT/mTOR;[13] and frequent mutations in the tumor suppressor *PTEN*.[15] Copy number aberrations affecting genes in the TCR signaling pathway are commonly observed in both variants.[15] Of these variants, the GATA3-PTCL group tends to be more uniform in immunophenotype (CD4$^+$/CD8$^-$) and gene expression, while the TBX21-PTCL group is more heterogeneous in immunophenotype and biomarker profile. Currently, the WHO considers the existing knowledge of the genetic landscape, clinical and pathologic correlations, and prognostic implications of these variants to be insufficient for formal assignment as PTCL-NOS

subtypes,[4] but efforts to refine these molecular variants and establish their translational relevance are ongoing.

T-cell development, activation, and differentiation

Much of the challenge of identifying the cell of origin of a T-cell malignancy lies in the complexity and plasticity of a T cell's life cycle, which spans multiple hematopoietic organs, transitions between numerous distinctive developmental stages, and culminates in a diverse array of terminal differentiation phenotypes (Figure 1).

Thymic ingress and T-lineage commitment

Early lymphoid precursors derived from hematopoietic stem cells are released in waves from the bone marrow into circulation and periodically are recruited to the thymus by chemokine and adhesion molecule signals from thymic microenvironmental niches.[16] These early progenitor cells have a high proliferative capacity and are characterized by surface expression of c-kit, Sca-1, Flt3, CD27, and variably IL7Ra.[16] Upon entry to the thymus, they are still capable of differentiating into other lymphoid cells, such as B cells and NK cells, as well as some myeloid cells, like dendritic cells. The loss of this multipotency and commitment to the T-cell lineage is facilitated by Notch signaling, which triggers expression of transcription factors T cell factor 1 (TCF-1), GATA3, and BCL11b that promote expression of genes required for T-cell survival and differentiation.[16–18]

Double negative thymocytes

These cells are then considered "double negative" (DN) thymocytes, a reference to their $CD4^{-}CD8^{-}$ immunophenotype. Four stages comprise the DN thymocytes—DN1 through DN4—distinguished by their variable expression of CD44 and CD25. Notch signaling drives the progression from $CD25^{-}CD44^{+}$ DN1 thymocytes to $CD25^{+}CD44^{+}$ DN2 thymocytes.[18] Between the DN2 and DN3 stage, thymocytes rearrange their TCR to become committed to either the α/β or

γ/δ T-cell lineage.[18] Rearrangement of the TCR, a process facilitated by enzymes encoded by *RAG1* and *RAG2*, allows for recognition of a multitude of peptide fragments by a host's T-cell repertoire.[17] The TCR exists on the surface of a T cell as heterodimers of either TCRα and TCRβ chains (the α/β TCR) or TCRγ and TCRδ chains (the γ/δ TCR). These heterodimers form complexes with surface CD3 molecules that conduct signals to various intracellular proteins through their cytoplasmic tails.[19]

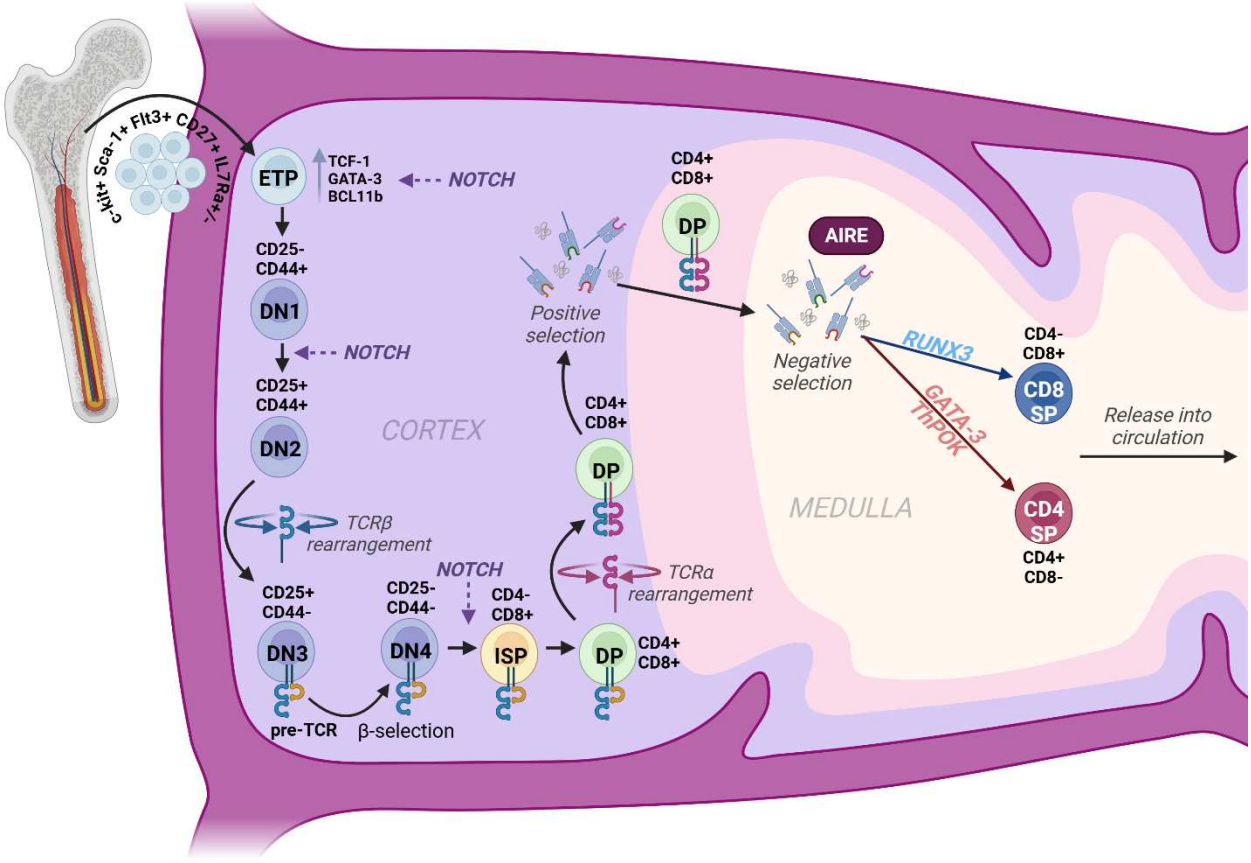


Figure 1: α/β T cell development in the thymus. Early lymphoid progenitors are released from the bone marrow in waves and periodically imported into the thymus, where they undergo a series of developmental stages to become mature $CD4^+$ or $CD8^+$ single positive (SP) thymocytes. Key milestones in thymocyte maturation include Notch signaling-induced T-cell lineage commitment in early thymic progenitor (ETP) cells, T-cell receptor (TCR) β chain rearrangement and β-selection in double negative (DN) thymocytes, TCR α chain rearrangement in double positive (DP) thymocytes, positive and negative selection of DP thymocytes with a complete α/β TCR, and CD4/CD8 lineage choice regulated by antagonistic transcription factors. Created in BioRender. Owens, E. (2025) https://BioRender.com/l98s694.

For an α/β TCR, the variable (V) and junctional (J) segments of the *TCRA* gene and V, J, and diversity (D) segments of the *TCRB* gene are rearranged. DNA breaks are then repaired after

rearrangement of these gene segments is complete.[17] The TCRβ chain is rearranged first, and prior to TCRα chain rearrangement, it associates with the invariant pre-Tα (pTα) chain and CD3 molecules to form the pre-TCR.[18,20] Only DN3 thymocytes that express a functional TCRβ chain are rescued from apoptosis and permitted to mature further, a checkpoint in thymocyte development known as *β-selection*. DN3 thymocytes during β-selection receive survival signals through their pre-TCR, and DN3 cells that pass β-selection will lose expression of CD25 and CD44 to become DN4 cells. DN4 cells then undergo a proliferative burst driven by pre-TCR and Notch signaling, maturing through a cycling $CD4^{-}CD8^{+}$ immature single positive (ISP) stage to the $CD4^{+}CD8^{+}$ "double positive" (DP) stage of thymocyte development.[20]

Double positive thymocytes

Rearrangement of the TCRα chain takes place during the DP stage of development, and this ultimately replaces the pTα chain of the pre-TCR.[20] Once DP thymocytes express the complete α/β TCR, their ability to recognize antigen presented by MHC molecules will be tested by cortical thymic epithelial cells presenting peptides complexed to self-MHC, a process known as *positive selection*. Those that successfully bind the peptide::self-MHC complex with moderate affinity are rescued from apoptosis and proceed to the thymic medulla for negative selection. During negative selection, the transcription factor AIRE drives the ectopic expression of various self-antigens by thymic medullary epithelial cells, and thymocytes whose TCRs exhibit reactivity for self-antigen are either destroyed or skewed toward a regulatory T cell lineage.[17,18]

CD4/CD8 lineage choice

Following positive and negative selection, DP thymocyte differentiation diverges again toward either the $CD4^{+}$ helper T cell lineage or $CD8^{+}$ cytotoxic T cell lineage. Interestingly, all DP thymocytes initially downregulate $CD8^{+}$ lineage-associated genes, becoming $CD4^{+}CD8^{lo}$

intermediate cells. Intermediate cells with weak TCR signals undergo coreceptor reversal induced by IL-7 signaling to upregulate the transcription factor *RUNX3* and commit the cell to a $CD8^+$ SP T cell. Intermediate cells with strong TCR signals, on the other hand, override these signals from IL-7, which instead upregulates transcription factors *GATA3* and *THPOK* to promote development into a $CD4^+$ SP T cell.[18,21]

Peripheral T-cell maturation

Once all these thymic developmental stages are completed, naïve α/β T cells are released into circulation and home to peripheral lymph nodes, where they are finally exposed to antigen and complete their differentiation into fully differentiated, activated T cells. T-cell activation requires two signals, one from the binding of the TCR by a peptide:MHC complex, and another costimulatory signal from binding of CD28 on T cells by B7 family proteins CD80 and CD86 on antigen presenting cells.[17,19] These binding events trigger a cascade of phosphorylation events that culminate in a T cell's secretion of interleukin-2 (IL-2) and subsequent clonal expansion, as well as activation of a variety of signaling pathways involved in immune cell survival, proliferation, and production of secreted proteins like cytokines that contribute to the immune response.[17,19] T-cell deactivation is accomplished by binding of inhibitory molecules cytotoxic T-lymphocyte-associated protein 4 (CTLA-4) and programmed cell death 1 (PD1) on T cells, as well as the actions of miscellaneous negative regulators of the various intracellular pathways activated by TCR signaling.[17,19]

Following activation, naïve $CD4^+$ T cells differentiate into a variety of mature phenotypes, including T helper 1 (Th1), T helper 2 (Th2), T helper 17 (Th17), T follicular helper (Tfh), T helper 9 (Th9), T helper 22 (Th22), or regulatory T cells (Treg), each able to uniquely produce and respond to certain cytokines and chemokines for a desired immune effect. Differentiation into these phenotypes is guided by specific master regulator transcription factors, and the resulting subtype

determines the T cell's function by influencing what cytokines and chemokines they produce and respond to.

Naïve T cells differentiate into Th1 cells in response to transcription factor T-bet (encoded by *TBX21*), which is activated by STAT4 and STAT1 signaling after exposure to IL-12 and IFN-γ released from macrophages and NK cells.[17] Th1 cells protect against intracellular pathogens like bacteria and viruses by producing IFN-γ and promoting class switching of B cells to IgG.[17]

Th2 cells are produced under the influence of the transcription factor GATA3. GATA3 can be upregulated in response to STAT6 activation by IL-4 secreted by dendritic cells and type 2 innate lymphoid cells, or independently of IL-4 via TCR signaling.[18,22] Th2 cells contribute to anthelmintic and allergic immune responses through secretion of IL-4, IL-5, IL-9, and IL-13, promotion of B-cell class switching to IgG and IgE, and induction of goblet cell hyperplasia in respiratory and gut epithelium.[17] They also contribute to tissue repair by polarizing macrophages toward a reparative and immunomodulatory M2 phenotype.[22]

Th17 cells are induced by transcription factor RORγt and protect against extracellular pathogens like bacteria and fungi through secretion of IL-17, IL-21, IL-22, and IL-23. TGF-β and IL-6 signaling help drive differentiation of Th17 cells by phosphorylating or sustaining activation of STAT3, which subsequently upregulates RORγt.[17,18]

Naïve $CD4^+$ T cells become Tfh cells under the influence of transcription factor BCL-6. These cells express CXCR5, which homes them to B-cell-rich follicles and drives the proliferation and affinity maturation of naïve B cells.[17,18]

Th9 cells are induced by transcription factors IRF4 and PU.1 and play a role in anthelmintic responses, allergy, and antitumor immunity through their production of IL-9, IL-10, and IL-21.[17,18] Upregulation of transcription factor IRF4 can result from IL-4 signaling and STAT6 phosphorylation,

while upregulation of PU.1 is often driven by TGF-β signaling.[18] Given some of their overlap with Th2 cells in function and modes of activation, it has been suggested that Th9 cells represent a subset of early activated Th2 cells.[17]

Th22 cells produce IL-22 and are induced by IL-23, IL-1, and IL-18 signaling.[17] Through their expression of IL-22, these cells help reduce the permeability of mucosal epithelial surfaces and regulate mucosal chemokine expression and the mucosal microbiome.[17] Unlike Th17 cells, which can also produce IL-22, Th22 cells do not secrete IL-17.[23] No Th22 cell-specific master regulator transcription factor has been identified; their differentiation and production of IL-22 appears to be dependent on aryl hydrocarbon receptor (AhR) and RORγt, but these transcription factors can also drive the differentiation and activities of Th17 cells.[17,23]

Treg cells are induced by transcription factor FOXP3, and their development in the periphery from conventional $CD4^+$ T cells is strongly influenced by IL-2 and TGF-β signaling.[17,18] Treg cells suppress immune responses to maintain immune homeostasis and promote tolerance for the prevention of autoimmunity. These cells have high expression of IL-2Ra (CD25) and secrete anti-inflammatory cytokines IL-10, TGF-β, and IL-35.[18] While some Tregs develop from conventional $CD4^+$ T cells in the periphery, the majority develop in the thymus. A portion of the DP thymocytes that express moderate levels of affinity for self-antigen during negative selection are retained to become Tregs, given their potential to suppress the host's own immune cells through their ability to recognize self-antigen.[17,18]

$CD8^+$ T cells exhibit less phenotypic diversity than their $CD4^+$ counterparts. Effector $CD8^+$ T cells are largely recognized for their cytotoxic activity against intracellular pathogens and cancer cells, accomplished by direct binding of Fas/Fas ligand, release of perforin granules that puncture target cell membranes to induce cell death, production of granzymes that activate a cascade of

caspases to induce target cell death, and secretion of IFN-γ and TNF-α to activate macrophages and enhance cytotoxicity.[17,18,24]

Most terminally differentiated effector T cells eventually contract after pathogen clearance and resolution of the immune response, but a smaller subset persists in circulation as long-lived memory T cells capable of reactivation upon repeat exposure to antigen. Both $CD4^+$ and $CD8^+$ T cells comprise the memory T cell repertoire, although $CD8^+$ T cells account for the majority.[24,25]

As mentioned previously, some thymocytes between the DN2 and DN3 stages will instead rearrange and express a γ/δ TCR. γ/δ T cells, which make up 1-5% of T cells in circulation and 20-50% of T cells at mucosal sites,[26] are unique in their ability to recognize peptides in a major histocompatibility complex (MHC)-independent manner, albeit with a more limited repertoire.[18] γ/δ T cells also acquire their effector function in the thymus,[18] unlike α/β T cells that undergo activation and terminal differentiation from a naïve phenotype in the periphery after exposure to antigen. The unique features of γ/δ T cells afford them innate-like functions, such as immune surveillance of tissues and modulation of local inflammation and tissue repair, in addition to their adaptive functions like immunologic memory.[18]

Cell of origin of PTCL-NOS

The cell of origin of human PTCL-NOS, the most common but least understood subtype of PTCL, remains a subject of ongoing investigation. The identification of the TBX21-PTCL and GATA3-PTCL subgroups led to postulation that these tumor cells are derived from their comparable normal Th1 and Th2 immune cell counterparts. However, GATA3 is upregulation is not restricted to Th2 lineage commitment; rather, this transcription factor is required for many stages of T-cell development, including the generation of early thymic T-lineage progenitors, the progression of thymocytes beyond the DN stage to the DP stage, and for the survival and maintenance of $CD4^+$ SP

thymocytes.[27,28] While target genes found in differentiated Th2 T-cell subsets (such as *CCR4*) are also upregulated in GATA3-PTCL, these are not Th2 specific. Recently, Geng et al.[29] demonstrated that *GATA3* is an oncogene highly expressed by multiple precursor-derived and mature T-cell lymphoma tumor types, as well as non-T-cell neoplasms like neuroblastoma and breast cancer, and cautioned against its expression being used to assume a Th2 ontology for PTCL-NOS. In this study, analysis of GATA3 binding revealed that only a small minority of the genes targeted by GATA3 in "mature" T-cell lymphomas consisted of Th2-associated genes, while the majority were genes involved in PI3K/AKT signaling or general T-cell growth and survival.

In mice, mature T cells were found to be incredibly resistant to oncogenic transformation by transduced retroviral vectors encoding potent T-cell oncogenes, suggesting that mature T-cell malignancies may actually arise from thymic precursors rather than mature T cells.[30] This has some support at the transcriptomic level in human PTCL-NOS, as markers of immaturity like *CD34* were overexpressed in tumor cells compared to control $CD4^+$ T cells in a recent gene expression profiling study of 68 cases of human PTCL-NOS[31] (determined by differential gene expression analyses performed with publicly available bulk RNA-sequencing data, GSE132550). While another gene expression profiling study comparing 28 cases of human PTCL-NOS with various purified T-cell subpopulations found the highest analogy to $CD4^+$ or $CD8^+$ activated peripheral T-cell gene signatures, this study notably did not include thymocytes in their analysis.[32]

Recurrent mutations in *TET2*, a gene normally highly expressed in hematopoietic stem cells and important for stem cell self-renewal and hematopoietic stem cell lineage commitment,[33] have been identified in 20-24% cases of human PTCL-NOS.[34,35] *TET2* inactivation in mice disrupts early and late stages of both myeloid and lymphoid hematopoiesis and gives cells a competitive advantage that eventually leads to malignant transformation, suggesting that *TET2* mutations predispose lymphomagenesis through impacts on the hematopoietic stem and precursor cell

compartments.[34] Thus, the prevalence of *TET2* mutations in human PTCL-NOS could indicate that some of the initiating mutations of this disease begin in early precursor immune cells.

There is some precedence for T-cell malignancies that originate in the thymus but present with a more mature phenotype. Historically, anaplastic large T-cell lymphoma (ALCL) was thought to arise from a mature, activated cytotoxic T cell given the evidence of clonal TCR rearrangement and production of cytotoxic molecules perforin and granzyme B by neoplastic ALCL cells.[36] Later, gene expression profiling studies identified overexpression of Th17-associated genes and production of IL-17 by neoplastic ALCL cells *in vitro*, shifting the presumed cell of origin to a mature, activated Th17 cell.[10,37] However, a more recent study by Malcolm et al.[38] using a transgenic *NPM-ALK* mouse model found that the initiation of ALCL malignancy begins in early DN thymocytes, where the driver *NPM-ALK* fusion gene mimics signaling of the pre-TCR to allow cells to bypass β-selection.

The well-established plasticity of T-helper cell differentiation further complicates this debate. Lineage tracing studies in mice have shown that mature T cells from various $CD4^+$ helper subtypes change their T-helper phenotype during their lifespan.[39] Modification of cell culture conditions can reprogram the subtype polarization of normal $CD4^+$ T cells and alter surface marker expression in human PTCL cells.[39,40] Taken together, these findings suggest that both normal $CD4^+$ T-cell and human PTCL phenotypes are not fixed.

Fusion genes as drivers of PTCL lymphomagenesis

Fusion genes describe a mutation in which two previously independent genes become juxtaposed, transcribed, and potentially translated together. The two main ways fusion genes drive oncogenesis are promoter exchange (e.g., the fusion of a promoter to a proto-oncogene) or the creation of chimeric genes that are translated as fusion proteins with oncogenic potential.[41] Because

of the inherent genomic instability of cancer cells, these rearrangements occur frequently, but the majority are "passenger" mutations with no functional relevance.[41] Those that do possess oncogenic properties are more anticipated to be driver mutations and tend to recur in multiple patients with the same tumor type.[41–44]

Fusion genes can arise from either chromosomal structural rearrangements like insertions, deletions, or translocations, or post-transcriptionally by fusion of mRNA transcripts through alternative splicing events.[43] Generally, fusion genes resulting from aberrations at the DNA level are more likely to be implicated in oncogenesis, while chimeric RNAs without evidence of DNA alteration are frequently detectable by PCR even in normal tissues, making their contribution to cancer more dubious.[45,46] Interestingly, in endometrial stromal tumors, a recurrent t(7;17)(p15;q21) chromosomal translocation results in the *JAZF1-SUZ12* fusion gene, but identical fusion RNA without the corresponding chromosomal translocation can also be detected in normal endometrial cells.[47] Similarly, the DNA translocation responsible for the *PAX3-FOXO1* fusion in rhabdomyosarcoma is not present in normal cells, but *PAX3-FOXO1* fusion RNA produced by trans-splicing is transiently seen in normal pluripotent cells as they differentiate into skeletal muscle.[48] These findings seem to support the importance of chromosomal evidence to validate the cancer specificity and oncogenic potential of a fusion gene. Nevertheless, cancer-promoting fusion RNAs without evidence of a concurrent chromosomal rearrangement have been demonstrated, such as the *SLC45A3-ELK4* fusion in prostate cancer and *DUS4L-BCAP29* in gastric cancer.[49–51]

Fusion genes have been implicated as driver mutations in human PTCL. In nTFHL and ATLL, a recurrent *CTLA4-CD28* fusion gene enhances CD28 signaling to promote T-cell activation and proliferation.[52–54] In the majority of ALK$^+$ ALCL, the constitutive activation of ALK tyrosine kinase and its aberrant expression in the nucleus and cytoplasm is caused by a t(2;5)(p23;q35) translocation that results in the creation of an NPM-ALK fusion protein.[55] A few recurrent fusion

genes described in ALK- ALCL can actually inform prognosis: 30-45% of cases have a rearrangement of 6p25.3 (involving either the *IRF4* or *DUSP22* genes in that region) with various partners, and rearrangements involving *DUSP22* are more often associated with good patient outcomes; 2-8% of ALK- ALCL have *TP63* rearrangements that are associated with worse outcomes than *DUSP22* rearrangements; and ALK- ALCL cases with no fusion genes have an intermediate prognosis.[56–58] In nTFHL and PTCL-NOS, the *ITK-SYK* fusion contributes to lymphomagenesis through constitutive activation of the TCR,[59–62] and the *FYN-TRAF3IP2* and *KHDRBS1-LCK* fusions constitutively activate NF-κB pathway signaling downstream of the TCR to drive oncogenesis.[63,64] Fusions involving the *VAV1* gene, which have been reported in up to 11% of PTCL-NOS, promote cell growth through increased Rac activation.[65]

In addition to the insight they offer into the pathogenesis of these tumors, fusion genes can represent a potential actionable target for novel personalized therapies. For example, treatment with the ALK inhibitor ceritinib has been efficacious in relapsed/refractory ALK^+ ALCL patients.[66] Blockade of CTLA4 can inhibit proliferation of *CTLA4-CD28* expressing cells *in vitro,*[52] and inhibitors of IκB kinase, a master regulator of the NF-κB signaling pathway, have been shown to be cytotoxic to *FYN-TRAF3IP2* expressing cells *in vitro* and capable of reducing tumor burden and improving survival in mice with *FYN-TRAF3IP2*-induced PTCL.[64] In one study,[61] mice transfected with the *ITK-SYK* fusion gene all developed T-cell lymphoproliferative disease except those concurrently treated with curcumin, a known inhibitor of SYK kinase activity, suggesting that SYK inhibitors may be of therapeutic interest to *ITK-SYK*$^+$ PTCLs.

Canine preclinical models of human cancer

In recent years, the domestic dog (*Canis familiaris*) has garnered interest as a model for a variety of human lymphomas given the resemblance of spontaneously arising canine hematopoietic malignancies to the clinical presentation, immunophenotypic features, molecular abnormalities, and

treatment responses of their human counterparts while also offering accelerated lifespans with shorter clinical courses of disease and less restrictive regulatory hurdles.[67,68] Because these are spontaneously arising tumors, they offer genetic diversity that is more comparable to that of human tumors, increased opportunities for investigation of the genetic origins, and greater insight into the tumor microenvironment when compared to traditional xenograft and genetically engineered mouse models.[67,69] While dogs are more outbred than these traditional mouse models, they still offer a degree of inbreeding with well-annotated, multi-generational pedigrees that allow for examination of breed-specific patterns of disease for the investigation of potential genetic predispositions.[67] Finally, pet dogs and their human owners share the same environment, which permits the consideration of environmental risk factors in the changing patterns of tumor development observed in both species.[67,70]

Canine peripheral T-cell lymphoma

Several studies have attempted to quantify subtypes of nodal lymphomas in dogs.[71–73] Unlike in people, ALCL and AITL appear to be very rare in dogs. Instead, the most common forms of nodal T-cell lymphoma in dogs include PTCL and T-zone lymphoma (TZL).[71] While TZL in humans is recognized solely as a morphologic variant of PTCL-NOS, TZL in dogs is typically discussed as a distinct entity due to its relatively high prevalence, its unique $CD3^+CD5^+CD25^+CD45^-$ immunophenotype, and its indolent clinical course.[74] Non-TZL PTCL in dogs, however, is an aggressive clinical disease with a more variable immunophenotype[75,76] and median survival times ranging from 97 to 235 days with CHOP or modified CHOP chemotherapy.[77,78]

Approximately 80% of non-TZL PTCLs in dogs exhibit a $CD3^+CD4^+CD45^+$ immunophenotype (referenced hereafter as $CD4^+$ PTCL), with frequent loss of surface CD5 expression.[75,79] Less frequently, canine PTCL may present with a $CD8^+CD4^-$ or $CD4^-CD8^-$ ("double negative") immunophenotype. Survival times between the different immunophenotypes are similar,

and $CD4^+$ PTCL and $CD4^-CD8^-$ PTCL tend to share similar clinical and flow cytometric features, such as frequent mediastinal involvement, hypercalcemia, and loss of surface CD5 expression.[76] In contrast, $CD8^+$ PTCL in dogs is unique for its frequent presentation with concurrent cutaneous lesions.[76]

The clinical, flow cytometric, cytologic, and histologic features of $CD4^+$ PTCL in dogs closely resemble PTCL-NOS.[75] Like in humans, peripheral nodal involvement is the most common presentation of canine $CD4^+$ PTCL,[80,81] males are overrepresented,[75,79] and it tends to present in advanced stages with short survival times and poor responses to CHOP-based chemotherapy.[77–80] Similar to the loss of surface CD5 expression frequently seen in canine $CD4^+$ PTCL, the majority of PTCL-NOS cases in humans lose expression of one or more mature T-cell antigens, most often CD5 or CD7.[82]

The potential utility of a canine model of PTCL-NOS is also exemplified by the relative frequency of its diagnosis in dogs compared to humans. In humans, the reported annual incidence of non-Hodgkin lymphoma is 19.6 per 100,000 people,[83] while in dogs, the crude annual incidence rate of canine lymphoma ranges from 63-107 per 100,000 dogs.[84–86] In people, PTCL accounts for 10-15% of all non-Hodgkin lymphoma, and ~30% of PTCLs belong to the PTCL-NOS subtype.[3] As mentioned previously, ALCL and AITL appear to be very rare in dogs, with most aggressive T cell lymphomas falling into the PTCL-NOS classification. Thus, not only do dogs appear to develop lymphoma more frequently, but PTCL-NOS is a relatively more commonly diagnosed subtype in dogs than it is in humans.

Investigations into the cell of origin of canine PTCL may also help elucidate the cell of origin of PTCL-NOS in humans. Canine $CD4^+$ PTCL frequently presents with a mediastinal mass and expresses low levels of CD25 and MHC class II.[75] In dogs, CD25 (IL2 receptor) is typically upregulated by T cells after they are activated.[87] In contrast with other species, all peripheral canine

T cells express class II MHC[88] and, consistent with observations in human T cells,[89] class II is believed to be upregulated upon T-cell activation. Thus, the low surface expression of CD25 and MHC class II in canine $CD4^+$ PTCL is more consistent with a cell of origin preceding the stage of T-cell activation.[75,90,91] As mentioned previously, the discovery of recurrent *TET2* mutations in subsets of human PTCL has prompted speculation about a progenitor cell of origin for these tumors, and mutations in *TET2* have been similarly reported in some cases of canine PTCL.[94] Additionally, 16-25% of canine PTCLs have been reported to have functional mutations in the *SATB1* gene,[75,95] a transcriptional repressor found predominantly in the nuclei of thymocytes that is necessary for the maturation of $CD4^+CD8^+$ thymocytes into single-positive cells.[96]

Conclusions

Despite some of the overlaps described above, similarities between the gene expression profile of human PTCL-NOS and canine PTCL remain largely unknown, and this gap in our understanding of how the molecular features of canine $CD4^+$ PTCL compare to human PTCL-NOS remain a hurdle to fully realizing the dog as a naturally occurring model of this poorly understood human disease. In a study of 6 cases of canine PTCL, surface expression of CD4 by flow cytometry was found to correlate with a distinctive histomorphology and gene expression profile.[75] While this study reported enrichment for PI3K/AKT/mTOR signaling and downregulation of PTEN in the transcriptome of canine $CD4^+$ PTCL (features which are also observed in the GATA3-PTCL variant of human PTCL-NOS), more comprehensive gene expression profiling studies are required to more rigorously establish the similarities between canine PTCL to human PTCL-NOS.

Additionally, there has been limited investigation into the cell of origin of canine PTCL using gene expression data. Some of the clinical and flow cytometric features of canine PTCL suggest that these tumors may arise from an earlier thymic precursor, a possibility that has been proposed for some human PTCLs previously thought to arise from mature T-cell subsets. Given other similarities

between canine PTCL and human PTCL-NOS, identifying the cell of origin of canine PTCL may offer additional insight into the disputed cell of origin of human PTCL-NOS.

While recurrent fusion genes have been implicated as driver mutations in small subsets of PTCL-NOS, these mutations have not been described in canine PTCL. Given other clinical and molecular similarities between canine $CD4^+$ PTCL and human PTCL-NOS, it is possible that gene fusions will be identified in canine PTCL that either resemble those identified in human PTCL-NOS or disrupt shared signaling pathways to similar oncogenic effect. Characterization of fusion genes in canine PTCL could therefore unveil opportunities for the further study of fusion genes as actionable drivers of human PTCL-NOS using a canine preclinical model.

To attempt to address these knowledge gaps, the objectives of this thesis are to 1) utilize gene expression profiling and variant analysis to determine whether the gene expression and mutational profile of canine PTCL is analogous to human PTCL-NOS, 2) investigate the cell of origin of canine PTCL using gene expression profiling and a single-cell transcriptomic atlas characterizing gene expression changes throughout normal canine T-cell development, and 3) further characterize the mutational landscape of canine PTCL by identifying recurrent fusion genes and other structural variants.

REFERENCES

1. Iqbal, J., Inghirami, G. & Chan, W. C. New insights into the biology of T-cell lymphomas. *Blood* 144, 1873–1886 (2024).

2. Stein, H. *et al.* CD30+ anaplastic large cell lymphoma: a review of its histopathologic, genetic, and clinical features. *Blood* 96, 3681–3695 (2000).

3. Yellala, A., Kallam, A. & Armitage, J. O. Geographic Distribution of the Peripheral T-cell Lymphomas. in *The Peripheral T-Cell Lymphomas* 57–68 (2021). doi:10.1002/9781119671336.ch5.

4. Alaggio, R. *et al.* The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. *Leukemia* 36, 1720–1748 (2022).

5. Campo, E. *et al.* The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood* 117, 5019–5032 (2011).

6. Harris, N. L. *et al.* A Revised European-American Classification of Lymphoid Neoplasms: A Proposal From the International Lymphoma Study Group. *Blood* 84, 1361–1392 (1994).

7. Matutes, E. The 2017 WHO update on mature T- and natural killer (NK) cell neoplasms. *Int. J. Lab. Hematol.* 40, 97–103 (2018).

8. Iqbal, J. & Leval, L. Pathology and Molecular Pathogenesis of T-Cell Lymphoma. in 95–141 (2019). doi:10.1007/978-3-030-00362-3_4.

9. Campo, E. *et al.* The International Consensus Classification of Mature Lymphoid Neoplasms: a report from the Clinical Advisory Committee. *Blood* 140, 1229–1253 (2022).

10. Iqbal, J. *et al.* Molecular signatures to improve diagnosis in peripheral T-cell lymphoma and prognostication in angioimmunoblastic T-cell lymphoma. *Blood* 115, 1026–1036 (2010).

11. de Leval, L. *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* 109, 4952–4963 (2007).

12. Bisig, B., Savage, K. J. & de Leval, L. Pathobiology of nodal peripheral T-cell lymphomas: current understanding and future directions. *Haematologica* 108, 3227–3243 (2023).

13. Iqbal, J. *et al.* Gene expression signatures delineate biological and prognostic subgroups in peripheral T-cell lymphoma. *Blood* 123, 2915–2923 (2014).

14. Ho, I.-C., Tai, T.-S. & Pai, S.-Y. GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation. *Nat. Rev. Immunol.* 9, 125–135 (2009).

15. Heavican, T. B. *et al.* Genetic drivers of oncogenic pathways in molecular subgroups of peripheral T-cell lymphoma. *Blood* 133, 1664–1676 (2019).

16. Wu, L. T lineage progenitors: the earliest steps en route to T lymphocytes. *Lymph. Dev. Tumour Immunol.* 18, 121–126 (2006).

17. Chopp, L., Redmond, C., O'Shea, J. J. & Schwartz, D. M. From thymus to tissues and tumors: A review of T-cell biology. *J. Allergy Clin. Immunol.* 151, 81–97 (2023).

18. Sun, L., Su, Y., Jiao, A., Wang, X. & Zhang, B. T cells in health and disease. *Signal Transduct. Target. Ther.* 8, 235 (2023).

19. Shah, K., Al-Haidari, A., Sun, J. & Kazi, J. U. T cell receptor (TCR) signaling in health and disease. *Signal Transduct. Target. Ther.* 6, 412 (2021).

20. Dutta, A., Zhao, B. & Love, P. E. New insights into TCR β-selection. *Trends Immunol.* 42, 735–750 (2021).

21. Brugnera, E. *et al.* Coreceptor Reversal in the Thymus: Signaled CD4+8+ Thymocytes Initially Terminate CD8 Transcription Even When Differentiating into CD8+ T Cells. *Immunity* 13, 59–71 (2000).

22. Walker, J. A. & McKenzie, A. N. J. TH2 cell development and function. *Nat. Rev. Immunol.* 18, 121–133 (2018).

23. Trifari, S., Kaplan, C. D., Tran, E. H., Crellin, N. K. & Spits, H. Identification of a human helper T cell population that has abundant production of interleukin 22 and is distinct from TH-17, TH1 and TH2 cells. *Nat. Immunol.* 10, 864–871 (2009).

24. Kumar, B. V., Connors, T. J. & Farber, D. L. Human T Cell Development, Localization, and Function throughout Life. *Immunity* 48, 202–213 (2018).

25. Künzli, M. & Masopust, D. CD4+ T cell memory. *Nat. Immunol.* 24, 903–914 (2023).

26. Silva-Santos, B., Serre, K. & Norell, H. γδ T cells in cancer. *Nat. Rev. Immunol.* 15, 683–691 (2015).

27. Pai, S.-Y. *et al.* Critical Roles for Transcription Factor GATA-3 in Thymocyte Development. *Immunity* 19, 863–875 (2003).

28. Hosoya, T. *et al.* GATA-3 is required for early T lineage progenitor development. *J. Exp. Med.* 206, 2987–3000 (2009).

29. Geng, X. *et al.* GATA-3 is a proto-oncogene in T-cell lymphoproliferative neoplasms. *Blood Cancer J.* 12, 149 (2022).

30. Newrzela, S. *et al.* Resistance of mature T cells to oncogene transformation. *Blood* 112, 2278–2286 (2008).

31. Etebari, M. *et al.* Transcriptional Analysis of Lennert Lymphoma Reveals a Unique Profile and Identifies Novel Therapeutic Targets. *Front. Genet.* 10, (2019).

32. Piccaluga, P. P. *et al.* Gene expression analysis of peripheral T cell lymphoma, unspecified, reveals distinct profiles and new potential therapeutic targets. *J. Clin. Invest.* 117, 823–834 (2007).

33. Solary, E., Bernard, O. A., Tefferi, A., Fuks, F. & Vainchenker, W. The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. *Leukemia* 28, 485–496 (2014).

34. Quivoron, C. *et al.* TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. *Cancer Cell* 20, 25–38 (2011).

35. Lemonnier, F. *et al.* Recurrent TET2 mutations in peripheral T-cell lymphomas correlate with TFH-like features and adverse clinical parameters. *Blood* 120, 1466–1469 (2012).

36. Foss, H.-D. *et al.* Anaplastic Large-Cell Lymphomas of T-Cell and Null-Cell Phenotype Express Cytotoxic Molecules. *Blood* 88, 4005–4011 (1996).

37. Schleussner, N. *et al.* The AP-1-BATF and -BATF3 module is essential for growth, survival and TH17/ILC3 skewing of anaplastic large cell lymphoma. *Leukemia* 32, 1994–2007 (2018).

38. Malcolm, T. I. M. *et al.* Anaplastic large cell lymphoma arises in thymocytes and requires transient TCR expression for thymic egress. *Nat. Commun.* 7, 10087 (2016).

39. DuPage, M. & Bluestone, J. A. Harnessing the plasticity of CD4+ T cells to treat immune-mediated disease. *Nat. Rev. Immunol.* 16, 149–163 (2016).

40. Mamand, S., Carr, M., Allchin, R. L., Ahearne, M. J. & Wagner, S. D. Interleukin-2-inducible T-cell kinase inhibitors modify functional polarization of human peripheral T-cell lymphoma cells. *Blood Adv.* 3, 705–710 (2019).

41. Ortiz de Mendíbil, I., Vizmanos, J. L. & Novo, F. J. Signatures of Selection in Fusion Transcripts Resulting From Chromosomal Translocations in Human Cancer. *PLOS ONE* 4, e4805 (2009).

42. Latysheva, N. S. & Babu, M. M. Discovering and understanding oncogenic gene fusions through data intensive computational approaches. *Nucleic Acids Res.* 44, 4487–4503 (2016).

43. Taniue, K. & Akimitsu, N. Fusion Genes and RNAs in Cancer Development. *Non-Coding RNA* 7, (2021).

44. Wang, X.-S. *et al.* An integrative approach to reveal driver gene fusions from paired-end sequencing data in cancer. *Nat. Biotechnol.* 27, 1005–1011 (2009).

45. Frenkel-Morgenstern, M. *et al.* Chimeras taking shape: potential functions of proteins encoded by chimeric RNA transcripts. *Genome Res.* 22, 1231–1242 (2012).

46. Babiceanu, M. *et al.* Recurrent chimeric fusion RNAs in non-cancer tissues and cells. *Nucleic Acids Res.* 44, 2859–2872 (2016).

47. Li, H., Wang, J., Mor, G. & Sklar, J. A Neoplastic Gene Fusion Mimics Trans-Splicing of RNAs in Normal Human Cells. *Science* 321, 1357–1361 (2008).

48. Yuan, H. *et al.* A Chimeric RNA Characteristic of Rhabdomyosarcoma in Normal Myogenesis Process. *Cancer Discov.* 3, 1394–1403 (2013).

49. Rickman, D. S. *et al.* SLC45A3-ELK4 Is a Novel and Frequent Erythroblast Transformation–Specific Fusion Transcript in Prostate Cancer. *Cancer Res.* 69, 2734–2738 (2009).

50. Qin, F., Zhang, Y., Liu, J. & Li, H. SLC45A3-ELK4 functions as a long non-coding chimeric RNA. *Cancer Lett.* 404, 53–61 (2017).

51. Kim, H.-P. *et al.* Novel fusion transcripts in human gastric cancer revealed by transcriptome analysis. *Oncogene* 33, 5434–5441 (2014).

52. Yoshida, N. *et al.* Genomic landscape of young ATLL patients identifies frequent targetable CD28 fusions. *Blood* 135, 1467–1471 (2020).

53. Hae Yong Yoo *et al.* Frequent CTLA4-CD28 gene fusion in diverse types of T-cell lymphoma. *Haematologica* 101, 757–763 (2016).

54. David Vallois *et al.* RNA fusions involving CD28 are rare in peripheral T-cell lymphomas and concentrate mainly in those derived from follicular helper T cells. *Haematologica* 103, e360–e363 (2018).

55. de Leval, L. & Gaulard, P. Tricky and terrible T-cell tumors: these are thrilling times for testing: molecular pathology of peripheral T-cell lymphomas. *Hematol. Am. Soc. Hematol. Educ. Program* 2011, 336–343 (2011).

56. Savage, K. J. & Slack, G. W. *DUSP22*-rearranged ALK-negative anaplastic large cell lymphoma is a pathogenetically distinct disease but can have variable clinical outcome. *Haematologica* 108, 1463–1467 (2023).

57. Sibon, D. *et al.* ALK-negative anaplastic large cell lymphoma with *DUSP22* rearrangement has distinctive disease characteristics with better progression-free survival: a LYSA study. *Haematologica* 108, 1590–1603 (2023).

58. Feldman, A. L. *et al.* Discovery of recurrent t(6;7)(p25.3;q32.3) translocations in ALK-negative anaplastic large cell lymphomas by massively parallel genomic sequencing. *Blood* 117, 915–919 (2011).

59. Huang, Y. *et al.* Peripheral T-cell Lymphomas With a Follicular Growth Pattern are Derived From Follicular Helper T Cells (TFH) and may Show Overlapping Features With Angioimmunoblastic T-cell Lymphomas. *Am. J. Surg. Pathol.* 33, (2009).

60. Streubel, B., Vinatzer, U., Willheim, M., Raderer, M. & Chott, A. Novel t(5;9)(q33;q22) fuses ITK to SYK in unspecified peripheral T-cell lymphoma. *Leukemia* 20, 313–318 (2006).

61. Dierks, C. *et al.* The ITK-SYK Fusion Oncogene Induces a T-Cell Lymphoproliferative Disease in Mice Mimicking Human Disease. *Cancer Res.* 70, 6193–6204 (2010).

62. Pechloff, K. *et al.* The fusion kinase ITK-SYK mimics a T cell receptor signal and drives oncogenesis in conditional mouse models of peripheral T cell lymphoma. *J. Exp. Med.* 207, 1031–1044 (2010).

63. Debackere, K. *et al.* Fusion transcripts FYN-TRAF3IP2 and KHDRBS1-LCK hijack T cell receptor signaling in peripheral T-cell lymphoma, not otherwise specified. *Nat. Commun.* 12, 3705 (2021).

64. Moon, C. S. *et al.* FYN–TRAF3IP2 induces NF-κB signaling-driven peripheral T-cell lymphoma. *Nat. Cancer* 2, 98–113 (2021).

65. Boddicker, R. L. *et al.* Integrated mate-pair and RNA sequencing identifies novel, targetable gene fusions in peripheral T-cell lymphoma. *Blood* 128, 1234–1245 (2016).

66. Gambacorti Passerini, C. *et al.* Crizotinib in Advanced, Chemoresistant Anaplastic Lymphoma Kinase–Positive Lymphoma Patients. *JNCI J. Natl. Cancer Inst.* 106, djt378 (2014).

67. Richards, K. L. & Suter, S. E. Man's best friend: what can pet dogs teach us about non-Hodgkin's lymphoma? *Immunol. Rev.* 263, 173–191 (2015).

68. Villarnovo, D., McCleary-Wheeler, A. L. & Richards, K. L. Barking up the right tree: advancing our understanding and treatment of lymphoma with a spontaneous canine model. *Curr. Opin. Hematol.* 24, 359–366 (2017).

69. Marconato, L., Gelain, M. E. & Comazzi, S. The dog as a possible animal model for human non-Hodgkin lymphoma: a review. *Hematol. Oncol.* 31, 1–9 (2013).

70. Gardner, H. L., Fenger, J. M. & London, C. A. Dogs as a Model for Cancer. *Annu. Rev. Anim. Biosci.* 4, 199–222 (2016).

71. Valli, V. E. *et al.* Classification of canine malignant lymphomas according to the World Health Organization criteria. *Vet. Pathol.* 48, 198–211 (2011).

72. Valli, V. E., Kass, P. H., Myint, M. S. & Scott, F. Canine Lymphomas: Association of Classification Type, Disease Stage, Tumor Subtype, Mitotic Rate, and Treatment With Survival. *Vet. Pathol.* 50, 738–748 (2013).

73. Ponce, F. *et al.* A Morphological Study of 608 Cases of Canine Malignant Lymphoma in France With a Focus on Comparative Similarities Between Canine and Human Lymphoma Morphology. *Vet. Pathol.* 47, 414–433 (2010).

74. Seelig, D. M. *et al.* Canine T-zone lymphoma: unique immunophenotypic features, outcome, and population characteristics. *J. Vet. Intern. Med.* 28, 878–886 (2014).

75. Harris, L. J. *et al.* Canine CD4+ T-cell lymphoma identified by flow cytometry exhibits a consistent histomorphology and gene expression profile. *Vet. Comp. Oncol.* 17, 253–264 (2019).

76. Harris, L. J. *et al.* Clinical features of canine nodal T-cell lymphomas classified as CD8+ or CD4−CD8− by flow cytometry. *Vet. Comp. Oncol.* 18, 416–427 (2020).

77. Rebhun, R. B. *et al.* CHOP chemotherapy for the treatment of canine multicentric T-cell lymphoma. *Vet. Comp. Oncol.* 9, 38–44 (2011).

78. Angelo, G., Cronin, K. & Keys, D. Comparison of combination l-asparaginase plus CHOP or modified MOPP treatment protocols in dogs with multi-centric T-cell or hypercalcaemic lymphoma. *J. Small Anim. Pract.* 60, 430–437 (2019).

79. Avery, P. R. *et al.* Flow Cytometric Characterization and Clinical Outcome of CD4+ T-Cell Lymphoma in Dogs: 67 Cases. *J. Vet. Intern. Med.* 28, 538–546 (2014).

80. Weisenburger, D. D. *et al.* Peripheral T-cell lymphoma, not otherwise specified: a report of 340 cases from the International Peripheral T-cell Lymphoma Project. *Blood* 117, 3402–3408 (2011).

81. Fournel-Fleury, C. *et al.* Canine T-cell Lymphomas: A Morphological, Immunological, and Clinical Study of 46 New Cases. *Vet. Pathol.* 39, 92–109 (2002).

82. Jha, K. K., Gupta, S. K., Saluja, H. & Subedi, N. Peripheral T-cell lymphoma, not otherwise specified. *J. Fam. Med. Prim. Care* 6, 427–430 (2017).

83. Howlader, N. *et al.* SEER Cancer Statistics Review, 1975-2017. *National Cancer Institute. Bethesda, MD* https://seer.cancer.gov/csr/1975_2017/ (2019).

84. Dobson, J. M., Samuel, S., Milstein, H., Rogers, K. & Wood, J. L. N. Canine neoplasia in the UK: estimates of incidence rates from a population of insured dogs. *J. Small Anim. Pract.* 43, 240–246 (2002).

85. Edwards, D. S., Henley, W. E., Harding, E. F., Dobson, J. M. & Wood, J. L. N. Breed incidence of lymphoma in a UK population of insured dogs. *Vet. Comp. Oncol.* 1, 200–206 (2003).

86. Pittaway, C., Schofield, I., Dobson, J., O'Neill, D. G. & Brodbelt, D. C. Incidence and risk factors for the diagnosis of lymphoma in dogs in UK primary-care practice. *J. Small Anim. Pract.* 60, 581–588 (2019).

87. Szopa, I. M. *et al.* Effective Activation and Expansion of Canine Lymphocytes Using a Novel Nano-Sized Magnetic Beads Approach. *Front. Immunol.* 12, 604066–604066 (2021).

88. Cobbold, S., Holmes, M. & Willett, B. The immunology of companion animals: reagents and therapeutic strategies with potential veterinary and human clinical applications. *Immunol. Today* 15, 347–353 (1994).

89. Oshima, S. & Eckels, D. D. Selective signal transduction through the CD3 or CD2 complex is required for class II MHC expression by human T cells. *J. Immunol.* 145, 4018 (1990).

90. Doveren, R. F., Buurman, W. A., Schutte, B., Groenewegen, G. & van der Linden, C. J. Class II antigens on canine T lymphocytes. *Tissue Antigens* 25, 255–265 (1985).

91. Saifuddin, M., Spear, G. T., Chang, C. & Roebuck, K. A. Expression of MHC class II in T cells is associated with increased HIV-1 expression. *Clin. Exp. Immunol.* 121, 324–331 (2000).

92. Cheng, S., Zhang, W., Inghirami, G. & Tam, W. Mutation analysis links angioimmunoblastic T-cell lymphoma to clonal hematopoiesis and smoking. *eLife* 10, e66395 (2021).

93. Yao, W.-Q. *et al.* Angioimmunoblastic T-cell lymphoma contains multiple clonal T-cell populations derived from a common TET2 mutant progenitor cell. *J. Pathol.* 250, 346–357 (2020).

94. McDonald, J. T. *et al.* Comparative oncology DNA sequencing of canine T cell lymphoma via human hotspot panel. *Oncotarget* 9, 22693–22702 (2018).

95. Elvers, I. *et al.* Exome sequencing of lymphomas from three dog breeds reveals somatic mutation patterns reflecting genetic background. *Genome Res.* 25, 1634–1645 (2015).

96. Alvarez, J. D. *et al.* The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. *Genes Dev.* 14, 521–535 (2000).

CHAPTER ONE: The gene expression and mutational profile of canine CD4$^+$ peripheral T-cell lymphoma[1]

Summary

Peripheral T-cell lymphoma (PTCL) refers to a heterogenous group of T-cell neoplasms with poor treatment responses and survival times. Canine PTCL clinically and immunophenotypically resembles the most common human subtype, PTCL-not otherwise specified (PTCL-NOS), leading to interest in this canine disease as a naturally occurring model for human PTCL. Gene expression profiling in human PTCL-NOS has helped characterize this ambiguous diagnosis into distinct subtypes, but similar gene expression profiling in canine PTCL is lacking. And while several recurrent gene mutations have been described in human PTCL-NOS, previous efforts to characterize the mutational landscape of canine PTCL have been limited by inadequate immunophenotyping to classify tumors or by small sample sizes. Addressing these knowledge gaps is a crucial step for validating that canine PTCL and human PTCL-NOS are molecularly similar diseases.

Two bulk RNA-sequencing (RNA-seq) experiments were performed on two cohorts of PTCL samples and sorted CD4$^+$ and CD8$^+$ lymphocytes from healthy control dogs. The first cohort, hereafter referred to as "Cohort 1," consisted of 33 dogs with either CD4$^+$ (26/33), CD8$^+$ (4/33), or CD4-CD8- (3/33) PTCL as diagnosed by flow cytometry and sorted CD4$^+$ and CD8$^+$ lymphocytes from 8 healthy control dogs. The second cohort, hereafter referred to as "Cohort 2," consisted of 96 dogs with CD4$^+$ PTCL diagnosed by flow cytometry, sorted CD4$^+$ nodal lymphocytes from 5 healthy control dogs, and sorted CD4$^+$ thymocytes from 2 healthy control dogs.

[1] Owens E, Harris L, Harris A, Yoshimoto J, Burnett R, Avery A. The gene expression profile and cell of origin of canine peripheral T-cell lymphoma. *BMC Cancer*. 2024 Jan 2;24(1):18. doi: 10.1186/s12885-023-11762-w.

Following normalization of RNA-seq data, we performed differential gene expression and unsupervised clustering methods. Gene set enrichment analysis (GSEA) was performed to determine the enrichment of canine CD4$^+$ PTCL for human PTCL-NOS and oncogenic signaling pathways. Variant analysis using CD4$^+$ PTCL samples from Cohort 2 and a panel of pooled normal canine samples was done to identify short variants, and these variant calls were correlated to previously reported recurrent gene mutations in canine T-cell lymphoma and human PTCL-NOS.

While the canine CD4$^+$ PTCL phenotype in Cohort 1 exhibited a consistent gene expression profile, the expression profiles of CD8$^+$ and CD4-CD8- canine PTCLs were more heterogeneous. In both cohorts, canine CD4$^+$ PTCL had increased expression of *GATA3* and other genes in the GATA3-PTCL expression profile, and enrichment for PI3K/AKT/mTOR signaling and signatures associated with downregulated PTEN activity, features consistent with the more aggressive GATA3-PTCL subtype of human PTCL-NOS. Several recurrently mutated genes in human PTCL-NOS were also mutated in canine CD4$^+$ PTCL, particularly those associated with the GATA3-PTCL subtype. In the case of *TP53* and *DNMT3A*, the protein domains implicated in the mutations were shared between canine CD4$^+$ PTCL and human PTCL-NOS. Additionally, some genes previously reported to be recurrently mutated in canine T-cell lymphoma were corroborated by our study, such as *SATB1* and *EGFR,* although many novel mutations not previously described in canine T-cell lymphoma were also identified. In conclusion, the gene expression and mutational profile of canine CD4$^+$ PTCL most closely resembled the GATA3-PTCL subtype of PTCL-NOS.

Introduction

Peripheral T-cell lymphoma (PTCL) refers to a poorly understood heterogeneous group of non-Hodgkin lymphomas arising from a T-lymphocyte cell of origin.[1] While these tumors are diverse in their clinical presentation and histologic and immunophenotypic features, they are similar in their aggressive clinical course, poor responses to treatment, and short survival times.[1] Given

their heterogeneity, attempts to subtype PTCLs based on their clinical and immunophenotypic features have been challenging, and approximately one-third of all PTCL cases fall under the "catch-all" diagnosis of PTCL-not otherwise specified (PTCL-NOS), which represents the most common type of PTCL diagnosed globally.[2]

Recently, gene expression profiling has helped further divide this ambiguous diagnosis into prognostically significant subtypes—TBX21-PTCL and GATA3-PTCL—with the GATA3-PTCL subtype being associated with lower overall survival times.[3] The TBX21-PTCL subtype is characterized by enrichment for features of T helper 1 (Th1) differentiation of helper T-cells, such as increased expression of IFN-γ and NF-κB, and enrichment for cytotoxic T-cell gene signatures.[3] The GATA3-PTCL subtype, on the other hand, is characterized by strong expression of the transcription factor GATA3,[3] which is essential for T-cell development and for promoting T helper type 2 (Th2) differentiation of helper T-cells;[4] enrichment for cellular proliferation pathways MYC and PI3K/AKT/mTOR;[3] and frequent mutations in the tumor suppressor PTEN.[5] Copy number aberrations affecting genes in the T-cell receptor signaling pathway are commonly observed in both subtypes.[5]

In dogs, the most common forms of nodal T-cell lymphoma include PTCL and T-zone lymphoma (TZL).[6] While TZL in humans is recognized solely as a morphologic variant of PTCL-NOS, TZL in dogs is typically discussed as a distinct entity due to its relatively high prevalence, its unique $CD3^+CD5^+CD25^+CD45$- immunophenotype, and its indolent clinical course.[7] Non-TZL PTCL in dogs, however, is an aggressive clinical disease with a more variable immunophenotype[8,9] and median survival times ranging from 97 days to 235 days with CHOP or modified CHOP chemotherapy.[10,11]

The clinical, flow cytometric, cytologic, and histologic features of non-TZL PTCL in dogs closely resemble PTCL-NOS,[8] and PTCL is relatively more common in dogs than in humans.[12] This

has led to their interest as a potential naturally occurring pre-clinical model for this disease. However, similarities between the gene expression profile of human PTCL-NOS and canine PTCL remain largely unknown. Currently, information on the gene expression profile of canine PTCL is limited to a single study of 6 cases of canine PTCL that found expression of CD4 by flow cytometry correlated with a distinctive histomorphology and gene expression profile.[8] While enrichment for the PI3K/AKT/mTOR signaling pathway and downregulation of *PTEN* was noted in these cases (features which are also observed in the GATA3-PTCL subtype of human PTCL-NOS), this study[8] did not make direct comparisons to human PTCL-NOS and did not report whether other features of the GATA3-PTCL or TBX21-PTCL subtypes of human PTCL-NOS were observed in their cohort.

Additionally, investigations into the mutational landscape of canine PTCL have previously been hindered by limited immunophenotyping and classification of these tumors. As mentioned previously, an important rule-out for nodal T-cell lymphoma in dogs is TZL, which has a distinctive immunophenotype and indolent clinical course distinguishing it from the more aggressive T-cell lymphomas that may capitulate human PTCL-NOS. Two of the existing studies attempting to characterize the mutational landscape of canine T-cell lymphoma limited their inclusion criteria to tumors positive for CD3 or CD5 by immunohistochemistry (IHC), or that had a clonal T-cell rearrangement by PCR for antigen receptor rearrangement (PARR), features that are observed with both PTCL and TZL in dogs. In McDonald et al.,[13] where T-cell lymphomas were diagnosed based on a CD5+CD79a- immunophenotype by IHC, DNA sequencing was performed on 16 canine T-cell lymphomas using a custom panel of human cancer mutation hotspots. The most commonly mutated genes identified in this study were *MET, KDR, STK11,* and *BRAF*, with fewer mutations detected in *SMAD4, TET2, PTEN, ATM, EGFR, JAK1, MYC, NOTCH1, SMO, TP53,* and *PLCG1*. In Elvers et al.,[14] T-cell lymphomas diagnosed based on a CD3+CD79a- immunophenotype by IHC or a clonal T-cell receptor rearrangement by PARR were further subdivided into 25 “Golden Retriever type” T-cell

lymphomas and 16 “Boxer type” T-cell lymphomas that had distinctive mutational profiles by whole exome sequencing. The most commonly mutated genes in “Golden Retriever type” T-cell lymphomas included *PSMA1, COX8A,* and *LTA4H*, while the most commonly mutated genes in “Boxer type” T-cell lymphoma included *SATB1* and *PTEN*—with both occurring in 25% of cases—followed by *MAP2K1, EEF1A1, NLRP14,* and *KCND2*.

A study by Harris et al.[8] performing variant analysis on bulk RNA-seq data from 6 canine PTCLs remains to our knowledge the only existing example of mutational profiling of canine PTCL where a non-TZL immunophenotype was confirmed by flow cytometry using a panel that included not only CD3 and CD5 but additional markers for CD25, CD45, CD4, and CD8 to definitively exclude TZL. In this study, one case of $CD4^+$ PTCL was heterozygous for a silent C>T single nucleotide polymorphism (SNP) at position 37 910 068 of the *PTEN* transcript, and one case of $CD4^+$ PTCL was heterozygous for a T>G SNP at position 24 651 976 in the *SATB1* gene transcript, resulting in a glutamine to arginine amino acid change. Notably, while mutations in *PTEN* and *SATB1* were described as mutations characteristic of the “Boxer type” lymphoma in the study by Elvers et al., these mutations in Harris et al. were identified in a Collie and Cavalier King Charles Spaniel and were not seen in PTCLs from the three Boxers in their study. While this study had the most rigorous immunophenotyping and classification of its T-cell lymphoma cases, it was the most limited of the three with regard to sample size. These deficiencies, either in appropriate classification of canine PTCL or in the number of dogs included, leave much of our understanding of the mutational landscape of canine PTCL still nebulous.

Therefore, the objectives of our study were to utilize bulk RNA-seq to 1) characterize the gene expression profile of larger cohorts of canine PTCL, 2) determine whether its gene expression profile is analogous to human PTCL-NOS to further establish the dog as a naturally occurring preclinical model of this disease, and 3) perform variant analysis on a larger cohort of definitive

non-TZL PTCLs to corroborate previously reported gene mutations in canine T-cell lymphoma and compare its mutational profile to that of human PTCL-NOS.

Materials and Methods

Sample collection

Fine needle tissue aspirates submitted to the Clinical Hematopathology Lab at Colorado State University for routine diagnostics and identified as PTCL by flow cytometry using the methods and antibody panels described in Seelig et al.[7] were considered for this study. The development and performance of these antibodies in dogs has been previously described.[15,16] Cases for Cohort 1 were diagnosed as PTCL by flow cytometry if they exhibited a homogeneous expansion of T cells, as characterized by the expression of pan-leukocyte antigen CD45, T-cell antigen CD3 (± CD5), and an absence of B-cell antigen CD21 or the stem cell antigen CD34. A previous study from our laboratory demonstrated that 70/70 cases of nodal lymphoma with this phenotype and histologic preparations sufficient for subtyping were consistent with the histologic classification of PTCL.[8] For Cohort 1, a total of 33 samples diagnosed as PTCL by these criteria were selected for RNA-sequencing (RNA-seq) analysis. The site sampled (as reported by the submitting veterinarians) was a peripheral lymph node in 29 cases, mediastinum in 3 cases, and bone marrow in 1 case (Supplemental Table 1). The selected cases exhibited variable expression of T-cell subset antigens CD4 and CD8 by flow cytometry and were subsequently categorized into $CD4^+$ PTCL (26/33 cases), $CD8^+$ PTCL (4/33 cases), and CD4-CD8- PTCL (3/33 cases). Cases for Cohort 2 were diagnosed as $CD4^+$ PTCL by flow cytometry if they exhibited a homogeneous expansion of T cells with a $CD45^+CD3^+CD4^+CD5^{+/-}CD21^-CD34^-$ immunophenotype. For Cohort 2, a total of 96 samples diagnosed with these criteria were selected for RNA-seq analysis. The sampled site was peripheral lymph node in 94 cases and mediastinum in 2 cases (Supplemental Table 1).

PTCL samples in both cohorts had a median viability of 90.83% (range 36.95-98.98%) and a median purity of 89.42% (range 26.5-98.26%) (Supplemental Table 2).

Control T cells for Cohort 1 were obtained from the lymph nodes of healthy 1-year-old male (2/9) and female (8/9) dogs designated for IACUC-approved surgical continuing education courses. Healthy control dogs were confirmed to have no evidence of lymphadenopathy, lymphoma, or other morbidities by physical examination, gross necropsy, and flow cytometry of the peripheral lymph nodes. Three of these healthy control dogs were Beagles, 5 were Hounds, and 1 was a mixed breed. Three of the control samples in our study were composed of pooled cells from 2 of the control dogs and the remaining 4 control samples consisted of cells from only one control dog (Supplemental Table 1). Cells were sorted into $CD4^+$ and $CD8^+$ T-cell subsets by a MoFlo cell sorter (Beckman Coulter, Brea, CA) using positive selection with anti-CD5 (clone YKIX322.3) and anti-CD4 (clone YKIX302.9) or anti-CD8 (clone YCATE55.9). Samples were only utilized if the desired population comprised over 90% of the sorted sample after purity analysis.

Control nodal T cells and thymocytes for Cohort 2 were collected from the lymph nodes of 7 healthy female mixed breed dogs ranging from 7 months to 1 year in age designated for IACUC-approved medical device research. Healthy control dogs were confirmed to have no evidence of lymphadenopathy, lymphoma, or other morbidities by physical examination and gross necropsy. Lymph node cells from all 7 dogs and thymocytes from 2 dogs were sorted into $CD4^+$ T-cell and $CD4^+$ SP thymocyte subsets, respectively, by a Sony SH800 cell sorter (Sony Biotechnology). Cells were sorted using positive selection with anti-CD5 (clone YKIX322.3) and anti-CD4 (clone YKIX302.9) and were only utilized if the desired population comprised over 90% of the sorted sample after purity analysis.

Flow cytometry of normal canine thymocytes

Normal thymocytes were obtained fresh from the thymus of 8 healthy female mixed breed dogs designated for IACUC-approved medical device research. Six of these dogs were 1 year old, 1 dog was 2 years old, and 1 dog was 7 months old (Supplemental Table 1). Surface CD4, CD8, CD25, and MHC class II expression on fresh thymocytes was evaluated by flow cytometry using the methods as described in Seelig et al.[7] and the flow cytometry antibody panels Panel 2 (for samples CF47 and CF48) and Panel 3 (for samples CF41-CF46) as described in Rout et al.[17]

RNA-sequencing

RNA from Cohort 1 PTCL and control samples was extracted using the Purelink RNA Mini Kit (Thermo Fisher Scientific, Waltham, MA) and quality measured with an Agilent 2100 Bioanalyzer System. The median RIN was 8.45, with a range of 6.3–9.8. RNA samples were delivered to Novogene Corporation Inc (Sacramento, CA) for library construction and sequencing. Sequencing was performed using an Illumina Hiseq PE150 platform. The number of raw reads, clean reads, raw bases, clean bases, and the error rate, phred quality scores, and GC content for each sample were reported by Novogene (Supplemental Table 3).

RNA from Cohort 2 $CD4^+$ PTCL and control samples was extracted using the Purelink RNA Mini Kit (Thermo Fisher Scientific, Waltham, MA) and quality measured with a NanoDrop One UV-Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, MA). The median A260/280 was 2.06, with a range of 1.96–2.2. The median A260/230 was 2.08, with a range of 1.76–2.32. RNA samples were delivered to Novogene Corporation Inc (Sacramento, CA) for library construction and sequencing. Sequencing was performed using an Illumina Hiseq PE150 platform. The number of raw reads, clean reads, raw bases, clean bases, and the error rate, phred quality scores, and GC content for each sample were reported by Novogene (Supplemental Table 3).

Data analysis

Adapters and low-quality reads were trimmed from raw RNA-seq FASTQ files using fastp[18] (version 0.23.1 for Cohort 1, version 0.23.3 for Cohort 2). Reads were then aligned to the CanFam3.1 reference genome (accessed through Ensembl release 104[19]) with STAR[20] (version 2.7.10a for Cohort 1 and version 2.7.10b for Cohort 2), and the number of reads per gene tabulated by featureCounts[21] (version 2.0.1 for Cohort 1 and version 2.0.3 for Cohort 2). For Cohort 2, reads were additionally aligned to the ROS_Cfam_1.0, CanFam4 (UU_Cfam_GSD_1.0), and CanFam6 (Dog10K_Boxer_Tasha) canine reference genomes, accessed through Ensembl release 109.[22] MultiQC (version 1.11 for Cohort 1 and version 1.14 for Cohort 2)[23] was used to obtain the alignment scores of raw reads. Data normalization using median of ratios, differential expression analysis, and principal component analysis (PCA) was conducted with DESeq2[24] (version 1.44.0) using default parameters. Low count genes (≤10) were pre-filtered prior to running DESeq2 functions. One sample in Cohort 1 (CF21) that had a variance in one PCA dimension that was >3x the standard deviation of all samples was excluded as an outlier from downstream differential expression analyses and gene set enrichment analyses. Volcano plots of gene expression data were generated with EnhancedVolcano[25] (version 1.22.0). Scatter and box plots of normalized count data were generated using ggplot2[26] (version 3.5.1), and means of normalized counts were compared by Wilcoxon signed-rank test using the ggpubr[27] (version 0.6.0) package.

Heatmaps of variance stabilized transformed normalized count data were generated and hierarchical clustering analyses performed using pheatmap (RRID:SCR_016418, version 1.0.12). Unsupervised hierarchical clustering by Euclidean distance with the Ward clustering method was performed on variance stabilized transformed normalized count data from the top 2000 genes with the highest median absolute derivation, and supervised hierarchical clustering by Euclidean distance with the Ward clustering method was performed on variance stabilized transformed

normalized count data from the list of genes associated with the GATA3-PTCL and TBX21-PTCL subtypes as defined in Iqbal et al.[3]

GSEA was performed with clusterProfiler[28] (version 4.12.6) and the Broad Institute's standard signal-to-noise and GSEAPreranked tools[29] (version 4.2.3) using the collective gene sets available through the Molecular Signatures Database (MSigDB).[30] The Wald test statistic was selected as the ranking metric for the list of differentially expressed genes in canine $CD4^+$ PTCL (as obtained by DESeq2) for its consideration of p-value in conjunction with log2 fold change. Default parameters were used for signal-to-noise GSEA analyses and the GSEAPreranked tool. clusterProfiler GSEA parameters were set to perform 10,000 permutations and enforce a p-value cutoff of 0.05 following correction using the Benjamini-Hochberg procedure.

The Broad Institute's GSEAPreranked tool (version 4.2.3)[29] was also used to compare the ranked list of differentially expressed genes in canine $CD4^+$ PTCL to the top 250 up- and downregulated genes in human PTCL-NOS, based on two previous microarray studies (GSE6338[31] and GSE132550[32]). The NCBI GEO2R differential gene expression analysis tool[33] was used to obtain the top 250 upregulated and downregulated genes (as ranked by the Wald test statistic) in human PTCL-NOS compared to control human $CD4^+$ lymphocytes from study GSE6338.[31] Non-normalized RNA-seq data from study GSE132550 was obtained through the NCBI Gene Expression Omnibus.[32] Data normalization and differential gene expression analysis was then performed using DESeq2 (version 1.34.0).[24] The full bioinformatics pipeline and software details for the bulk RNA-seq analysis have been made available in a public Github repository: https://github.com/pirateyoho/CaninePTCLRNASeq.

A Tukey's multiple comparisons test was performed using GraphPad Prism for Windows, GraphPad Software, San Diego, California USA, www.graphpad.com (version 10.0.2) to evaluate

differences in surface marker expression measured by flow cytometry between canine thymocytes, canine PTCL, and normal canine nodal $CD4^+$ T-cells.

Variant analysis using bulk RNA-seq reads from Cohort 2 was performed according to the Broad Institute's Genome Analysis Toolkit (GATK) Best Practices™.[34] Raw fastq files of whole exome sequencing (WES) data from 37 samples of normal canine blood provided by our collaborator Dr. Adam Harris were used to generate a panel of normals. Raw WES reads were trimmed with TrimGalore[35] (version 0.6.10) and aligned to the CanFam3.1 reference genome with Burrows-Wheeler Aligner (BWA)[36] (version 0.7.17). GATK (version 4.4.0.0) tools were used to mark duplicated reads, recalibrate base quality scores, and call short variants in normal samples to generate a somatic panel of normals. Binary alignment map (BAM) files generated by STAR (version 2.7.10b) aligning bulk RNA-seq reads to the CanFam3.1 reference genome were processed with the GATK SplitNCigarReads, which handles splicing events in RNA-seq data prior to variant prediction by splitting reads into exon segments and clipping any sequences overhanging into intronic regions. Standard GATK tools were then used to recalibrate base quality scores and call short variants in tumor samples, using the panel of normals and two other reference panels of known canine germline variants were also provided to help distinguish true variants from false positives. Single nucleotide polymorphism (SNP) and insertion-deletion (INDEL) calls were then filtered according to standard hard filtering recommendations in GATK Best Practices™ (Table 1).

Table 1. Criteria for hard-filtering short variants. Variant calls in RNA-seq data from Cohort 2 $CD4^+$ PTCLs that did not meet these thresholds were excluded from downstream analysis.

Filter	SNPs	INDELs
Unfiltered depth of variant samples (QD)	<2	<2
Strand bias estimated by the symmetric odds ratio test (SOR)	>3	N/A
Strand bias estimated using Fisher's exact test (FS)	>60	>200
Root mean square of the mapping quality of reads across all samples (MQ)	<40	N/A

Rank sum test for mapping qualities of reference (REF) versus alternate (ALT) allele reads (MQRankSum)	< -12.5	N/A
Rank sum test for relative positioning of REF versus ALT alleles within reads (ReadPosRankSum)	< -8	< -20

Filtered variant call files (VCF) were then annotated with Ensembl Variant Effect Predictor[37] (release 111) and converted to the Mutation Annotation Format (MAF) with vcf2maf[38] (version 2.4). Intronic variants and those predicted to have only low or modifier effects were filtered out and excluded from downstream analysis. Variants annotated as splice site or splice region were also ultimately excluded, given their predominance on initial evaluation (Supplemental Figure 1) and known challenges with false positive splice variant calls when using RNA-seq data.[39] The final MAF file versions were summarized, analyzed, and visualized with maftools[40] (version 2.20.0). After initial analysis in maftools, shared variants present in both the MAF file for canine $CD4^+$ PTCL samples and the MAF file containing variant calls for control canine $CD4^+$ nodal lymphocytes in Cohort 2 were filtered out from the tumor MAF file, and subsequent downstream analyses were performed with only $CD4^+$ PTCL-specific variants. Shared variants were defined as those spanning the same start position and end position on the same chromosome and assigned the same variant classification (nonsense, frameshift, etc.) and variant type (SNP, INDEL). The OncoKB™ Cancer Gene List[41,42] was used to annotate cancer-associated genes. Mutated oncogenic pathways were identified by supplying the *pathways = 'sigpw'* argument to the oncoplot() function in maftools, which derives its pathway gene lists from Sanchez-Vega et al (doi: 10.1016/j.cell.2018.03.035). Lollipop plots depicting the distribution of variant calls along protein sequences were generated using lollipops (version 1.7.2) using protein sequence isoforms extracted from UniProt[43] using queryup (version 1.0.5). UniProt isoforms corresponding to the CanFam3.1 Ensembl transcript ID for each gene were selected. If the associated UniProt entry was inactive, the Basic Local Alignment Search Tool (BLAST) was used to identify the active UniProt isoform with the highest sequence

similarity. The full bioinformatics pipeline and software details for variant analysis has been made available in a Github repository: https://github.com/pirateyoho/GATKVariantAnalysisRNASeq.

Results

Clinical features of canine PTCL

The most represented breeds across both cohorts included mixed breed dogs (35/129, 27%), Boxers (29/129, 23%), and Golden Retrievers (10/129, 8%) (Figure 2). Patient ages ranged from 2 years to 13 years, with a median age of 6 years. Male dogs were overrepresented at 84/129 cases (65%). These findings are consistent with previous demographic data.[8,44]

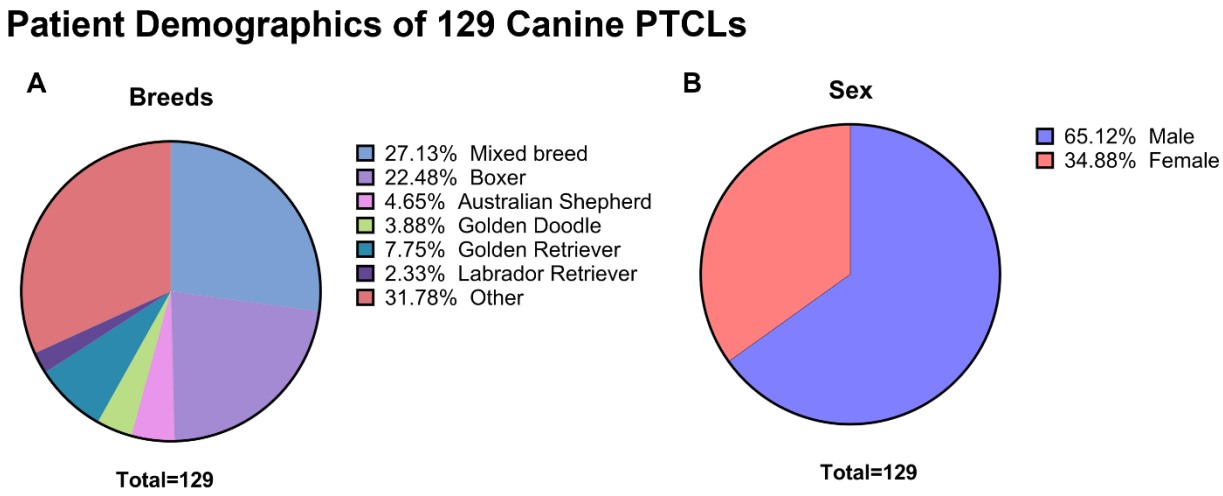


Figure 2: Patient demographics of canine PTCL. Patient demographics of canine PTCL. Across Cohorts 1 and 2, mixed breed dogs, Boxers, and Golden Retrievers were the most represented breeds (**A**), and male dogs were overrepresented (**B**).

Clinical features for all samples are reported in Supplemental Table 4. The most frequent clinical presentation of canine PTCL across both cohorts was peripheral lymphadenopathy, which was reported in 120/129 cases (93%). For the 52 cases where thoracic imaging was performed, a mediastinal mass was reported in 21 (40%) cases. For the 99 cases where standard blood chemistry was performed, hypercalcemia was reported in 47 (48%). For cases with abdominal imaging, 19/53 (36%) and 12/47 (26%) had reported splenomegaly and hepatomegaly, respectively. In one case

with splenomegaly (CF20), cytology attributed the cause to extramedullary hematopoiesis. Other more sporadically reported clinical signs included visceral lymphadenopathy (21/129 cases, 16%), neutrophilia (9/129 cases, 7%), pleural effusion (5/129 cases, 4%), and thrombocytopenia (24/129 cases, 19%) (Figure 3). One case (CF33) had a concurrent cutaneous melanoma diagnosed by cytology, one case (CF23) had a reported history of blindness and possible brain tumor, and an intestinal mass was reported in one case (CF137). Information on bone marrow involvement in this study is limited to 1 case (in which $CD4^+$ PTCL was diagnosed in a sample of bone marrow), as this site is not routinely sampled in dogs.

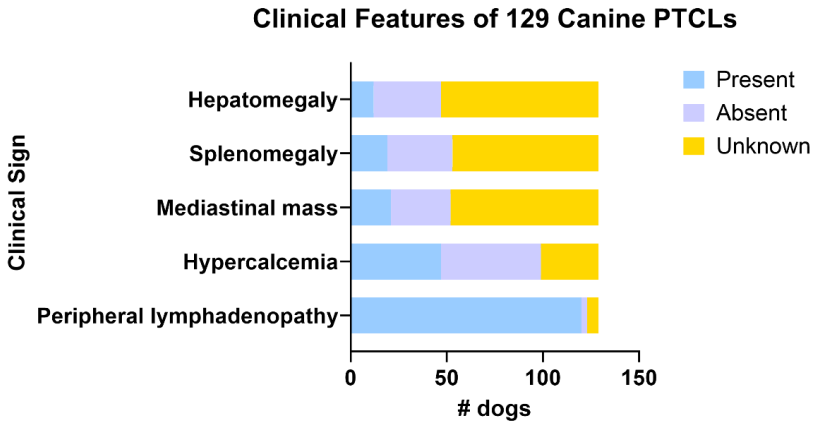


Figure 3: Clinical signs associated with canine PTCL. Across Cohorts 1 and 2, peripheral lymphadenopathy was the most commonly reported clinical sign. For cases where imaging was available, mediastinal masses were reported in 40% of cases, and hypercalcemia was reported in nearly half of cases where a biochemistry panel was performed.

Comparing canine reference genomes

For Cohort 2, RNA-seq reads were aligned to all four of the major canine reference genomes to compare mapping rates for each. The median percentage of reads able to be mapped to genes was 66%, 71%, 75.1%, and 70.7% for the CanFam3.1, ROS Cfam 1.0, CanFam4 (UU_Cfam_GSD_1.0) and CanFam6 (Dog10K_Boxer_Tasha) genomes, respectively (Figure 4A). However, the annotation of these genes is an equally important consideration when choosing a reference genome for RNA-seq analysis, as many downstream applications rely on association of mapped genes with HUGO Gene Nomenclature (HNGC) gene symbols. Annotations for a genome

are provided as general feature format (GFF) or general transfer format (GTF) files. For the CanFam3.1 genome, the GTF file (accessed through Ensembl release 104) describes 30,951 total features, 20,199 of which are annotated as protein-coding genes (65.3%). Of these, 15,824 have associated HNGC symbols (78.3%). The GTF file for the ROS Cfam 1.0 genome (accessed through Ensembl release 109) contains 31,121 total features, 20,500 of which are protein-coding (65.8%), and 16,028 of those protein-coding genes are associated with HGNC symbols (78.1%). The CanFam4 GTF file (accessed through Ensembl release 109) describes 31,398 total features, with 20,974 being protein-coding (66.8%). Of these protein-coding genes, 10,759 are annotated with corresponding HNGC gene symbols (51.3%). Finally, the GTF file for CanFam6 (accessed through Ensembl release 109) lists 29,993 features, of which 19,803 are protein-coding (66.0%), and 14,355 (72.5%) of these protein-coding genes are correlated with an HNGC gene symbol (Figure 4B). Thus, while the highest percentage of RNA-seq reads were able to be mapped to the CanFam4 reference genome, we elected to use CanFam3.1-aligned data for downstream analysis given its more robust gene annotation and consistency with the Cohort 1 study.

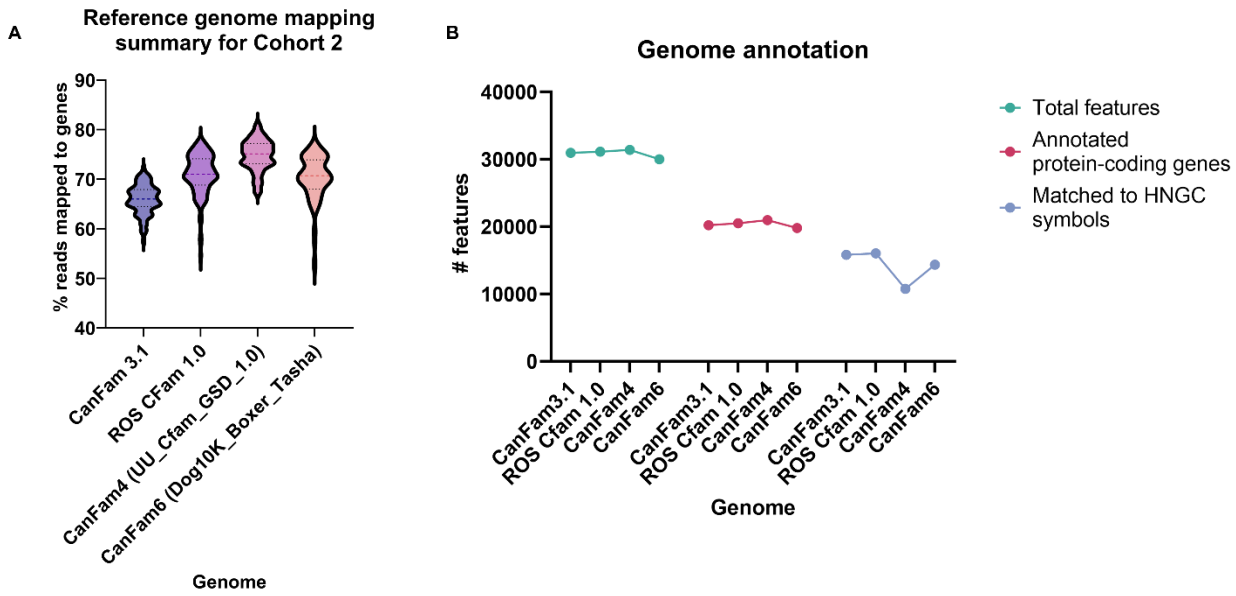


Figure 4: A comparison of read mapping rates and annotation across 4 canine reference genomes. While the greatest proportion of reads were able to map to genes using the CanFam4 reference genome (**A**), gene annotation for downstream analysis is superior in the CanFam3.1 and ROS Cfam 1.0 genomes (**B**).

CD4 expression by flow cytometry identifies a population of canine T-cell lymphomas with a unique gene expression profile

By flow cytometry, cases of canine PTCL in Cohort 1 varied in their expression of CD4 and CD8, allowing subgrouping into the following most common immunophenotypes: $CD4^+$ PTCL, $CD8^+$ PTCL, and CD4-CD8- PTCL. Among the $CD4^+$ PTCL cases in Cohort 1, loss of CD5 expression was observed by flow cytometry in 10/26 (38%) cases. 1 case categorized as $CD4^+$ PTCL demonstrated loss of CD3 expression by flow cytometry. In all but 1 case, CD25 expression was detected in < 1% of neoplastic cells by flow cytometry (Supplemental Table 2), and all tumors demonstrated CD45 expression by flow cytometry, ruling out TZL as a potential diagnosis. Among the PTCL cases in Cohort 2, which included only the $CD4^+$ PTCL phenotype, loss of CD5 was appreciated by flow cytometry in 42/96 (44%) cases, 1 case had loss of CD3 expression, and all tumors demonstrated CD45 expression by flow cytometry. All PTCL immunophenotypes in both cohorts exhibited significantly lower expression of MHC class II by flow cytometry compared to mature $CD4^+$ T-cells from the lymph nodes of healthy control dogs, more closely resembling normal thymocytes (Figure 5, Supplemental Table 2, Supplemental Table 5, Supplemental Table 6).

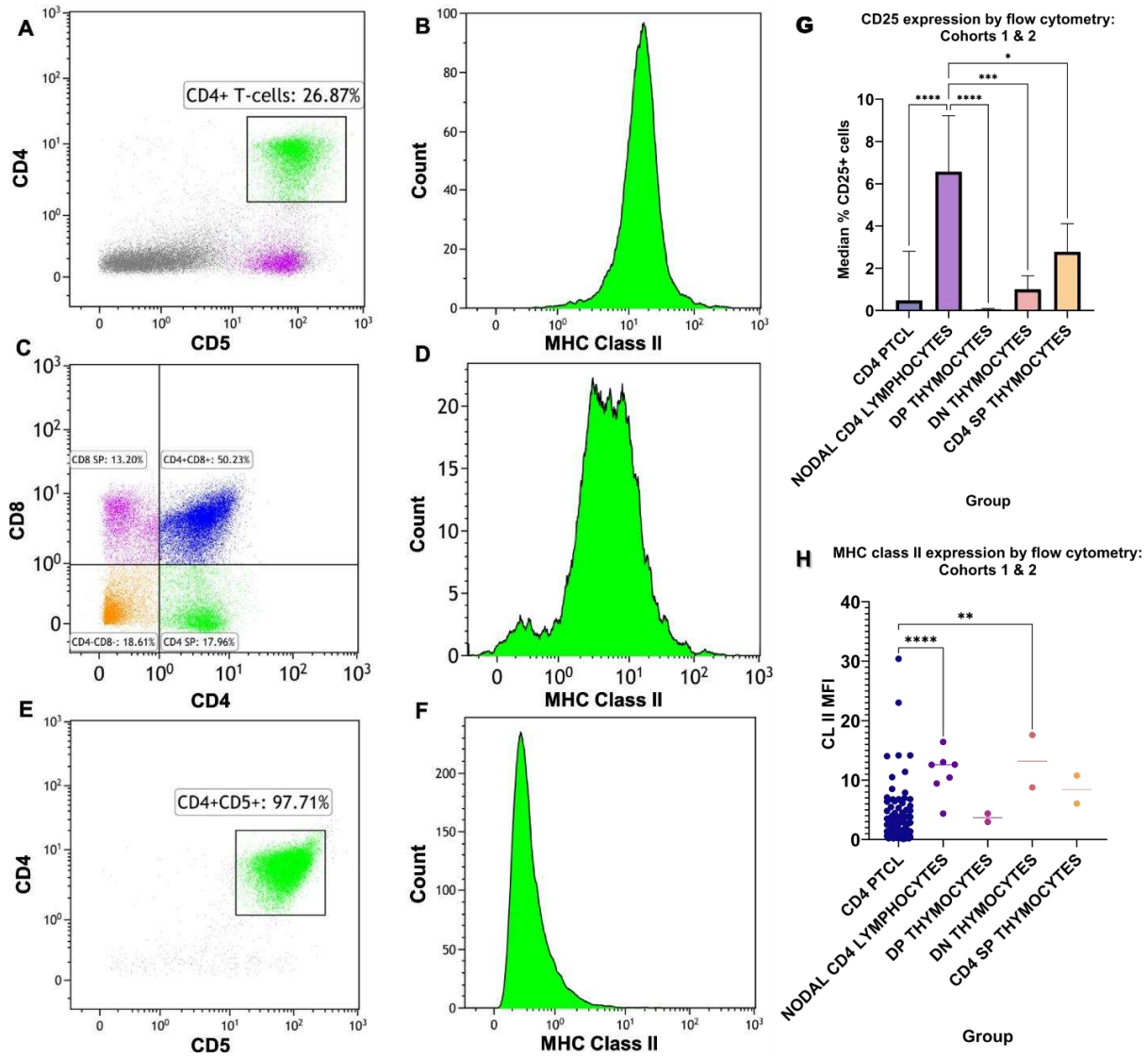


Figure 5: Flow cytometric features of normal mature CD4+ T-cells and normal CD4 SP thymocytes compared to canine CD4+ PTCL cells. (**A**) Flow cytometry of a normal lymph node from a healthy control dog. CD4+ CD5+ T-cells are highlighted in green. (**B, H**) Normal mature nodal CD4+ T-cells exhibit high MHC class II expression. (**C**) Flow cytometry of normal thymocytes from a healthy control dog. CD4 SP thymocytes are highlighted in green. (**D**) CD4 SP thymocytes exhibit lower MHC class II expression compared mature nodal CD4+ T-cells. (**E**) Flow cytometry of canine CD4+ PTCL reveals a homogeneous population of CD4+ CD5+ T-cells. (**F, H**) Canine CD4+ PTCL cells exhibit lower MHC class II expression compared to nodal mature T-cells from healthy control dogs. (**G**) Canine CD4+ PTCL cells and thymocytes exhibit lower CD25 expression compared to nodal mature T-cells from healthy control dogs.

Initial principal component analysis (PCA) of differential gene expression data from Cohort 1 revealed one outlier case (CF21, Supplemental Figure 2) in the CD4^{+} PTCL group that was subsequently excluded from subsequent PCA and downstream analyses. Notably, this case had high expression of CD30 (normalized count = 8471.48, mean = 358.98, standard deviation = 1622.66), a feature of anaplastic large cell lymphoma[45] which may have accounted for its clustering away from

the typical canine CD4+ PTCL cases. PCA revealed consistent clustering of control canine CD4 and CD8 lymphocytes away from tumor samples, and close clustering of canine CD4+ PTCLs (Figure 6A), corroborating previous findings that CD4 expression by flow cytometry corresponds to a distinct subgroup of canine PTCL with a uniform gene expression profile.[8] With one exception, the CD8+ and CD4-CD8- tumors did not cluster with the CD4+ lymphomas and exhibited more heterogeneity, suggesting these tumor types are from more than one lineage, and from a different lineage than CD4+ PTCL (Figure 6A). In the larger Cohort 2 study where the CD4+ immunophenotype was the only one represented, CD4+ PTCLs remained uniformly clustered without the formation of discrete subgroups (Figure 6B). We elected to focus on the distinctive canine CD4+ PTCL immunophenotype for the remainder of this thesis.

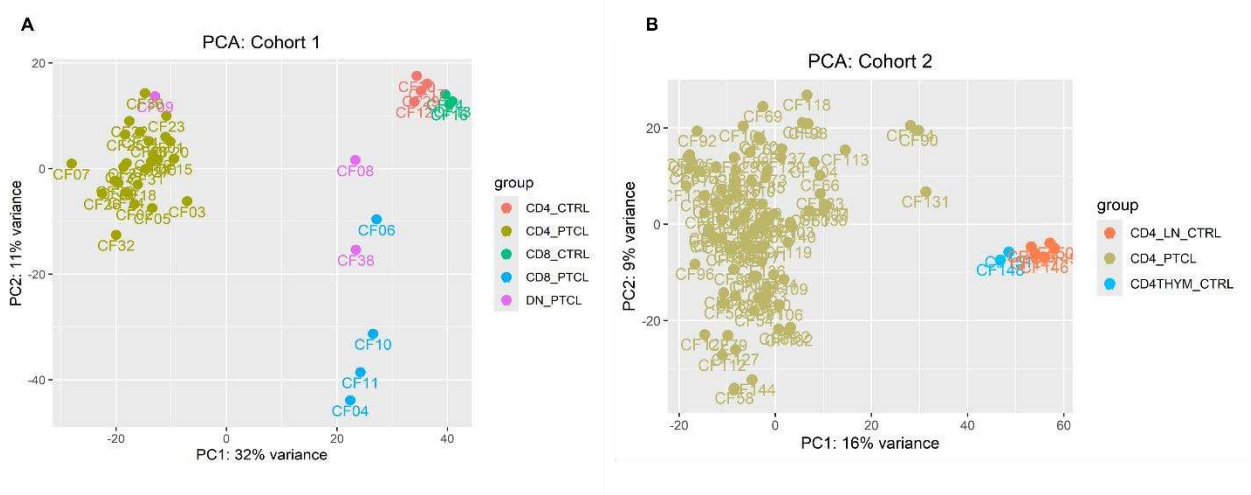
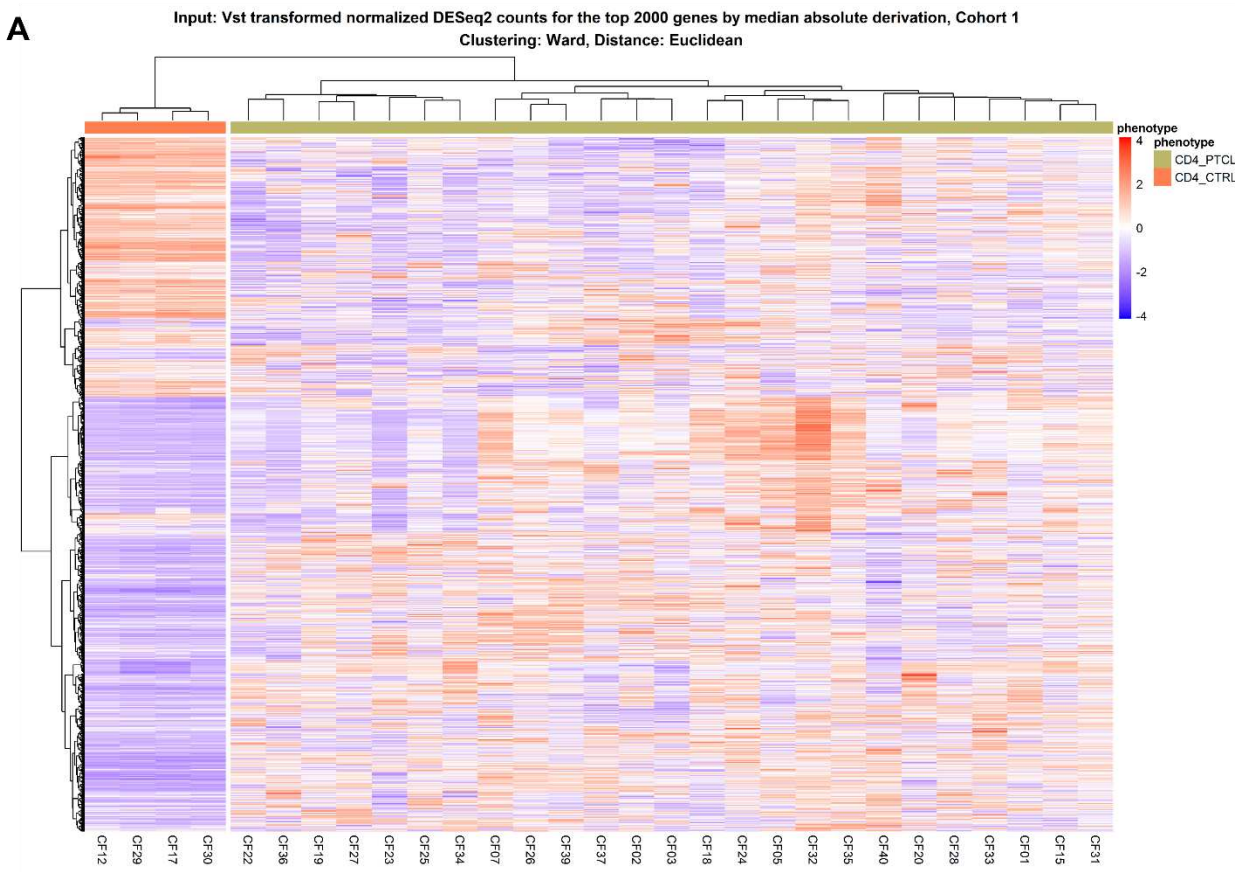


Figure 6: Principal component analysis (PCA) of canine PTCL and control samples. (**A**) Cohort 1. Control CD4+ lymphocytes (red) and control CD8+ lymphocytes (green) consistently clustered away from tumor samples. CD4 expression by flow cytometry identified a distinctive subtype of canine PTCL with a uniform gene expression profile (yellow). Wider variance was observed between CD8+ PTCL and CD4-CD8- PTCL samples, suggesting a more heterogeneous gene expression profile among these canine PTCL immunophenotypes. (**B**) Cohort 2. In a study of exclusively CD4+ PTCL, control CD4+ lymphocytes (red) and CD4+ thymocytes (blue) consistently cluster away from tumor samples. The global gene expression profile of canine CD4+ PTCL is relatively uniform.

<u>Canine CD4+ PTCL consistently exhibits a gene expression profile that is distinct from that of control CD4+ lymphocytes</u>

Quality parameters for the RNA-seq run are summarized in Supplemental Table 3. Alignment and assignment rates to the CanFam3.1, ROS Cfam 1.0, CanFam4, and CanFam6 reference genomes are summarized in Supplemental Table 7, Supplemental Table 8, Supplemental Table 9, and Supplemental Table 10, respectively. The median number of raw reads per sample was 58,089,233 (range: 46,655,240–81,854,562 reads) for Cohort 1 and 119,450,768 (range: 99,354,390–158,564,728) for Cohort 2. Error rates for both Cohort 1 and Cohort 2 ranged from 0.02% to 0.03%. The average GC content was 50.88% for Cohort 1 and 49.48% for Cohort 2. For both cohorts, the average percentage of reads able to be uniquely mapped to the CanFam3.1 reference genome per sample was 93.74%, and the average percentage of reads that were found to overlap known genes was 61.07%.

Unsupervised hierarchical clustering of variance stabilized transformed count data from $CD4^+$ PTCL samples and control $CD4^+$ lymphocytes using average Euclidean distance and Ward clustering of the top 2000 genes with the highest median absolute derivation revealed a consistent gene expression profile across all control canine $CD4^+$ lymphocyte samples from healthy dogs (Figure 7). All canine $CD4^+$ PTCL samples demonstrated a more heterogeneous gene expression profile that was distinct from that of control $CD4^+$ lymphocytes (Figure 7).


A
Input: Vst transformed normalized DESeq2 counts for the top 2000 genes by median absolute derivation, Cohort 1
Clustering: Ward, Distance: Euclidean
phenotype
phenotype
CD4_PTCL
CD4_CTRL
4
2
0
-2
-4
CF12
CF29
CF17
CF30
CF22
CF36
CF19
CF27
CF23
CF25
CF34
CF07
CF26
CF39
CF37
CF02
CF03
CF18
CF24
CF05
CF32
CF35
CF40
CF20
CF28
CF33
CF01
CF15
CF31

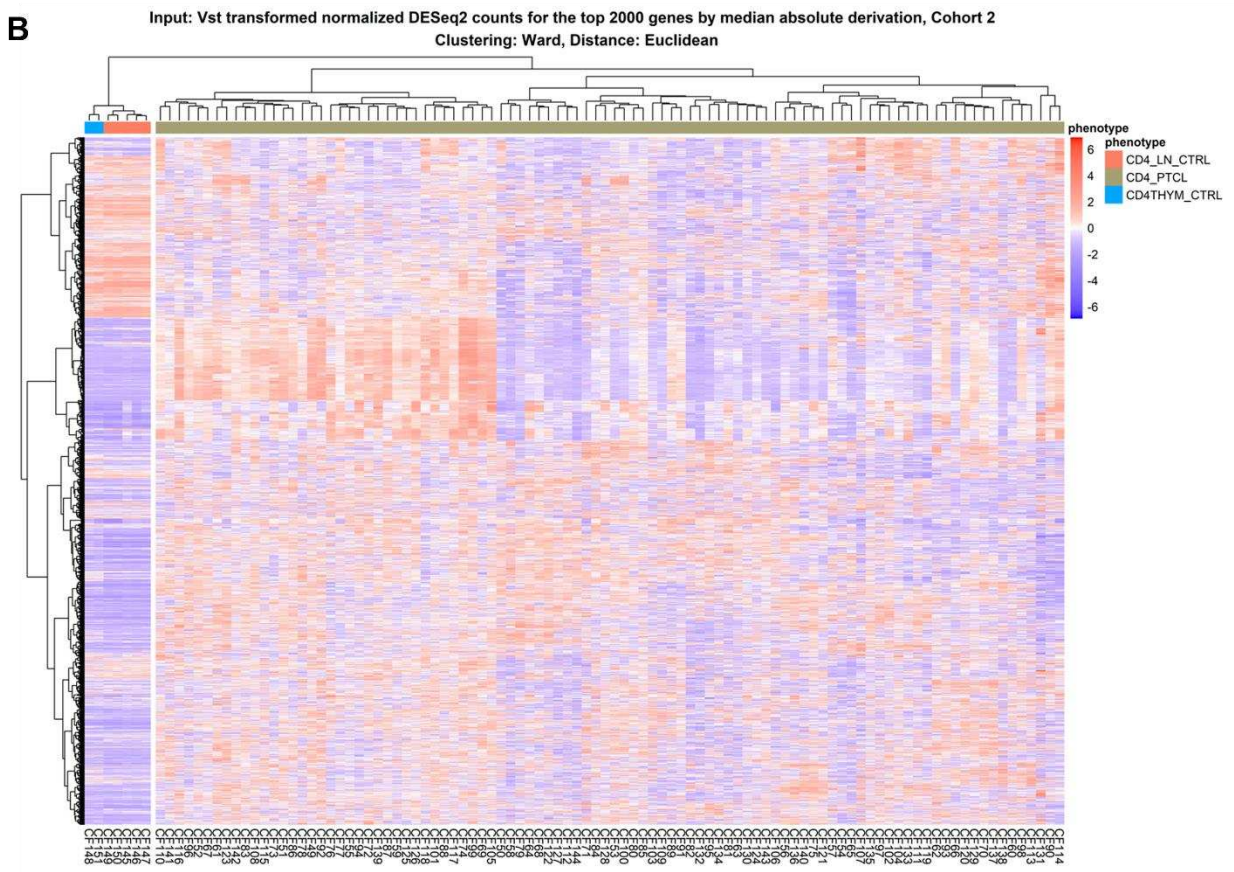
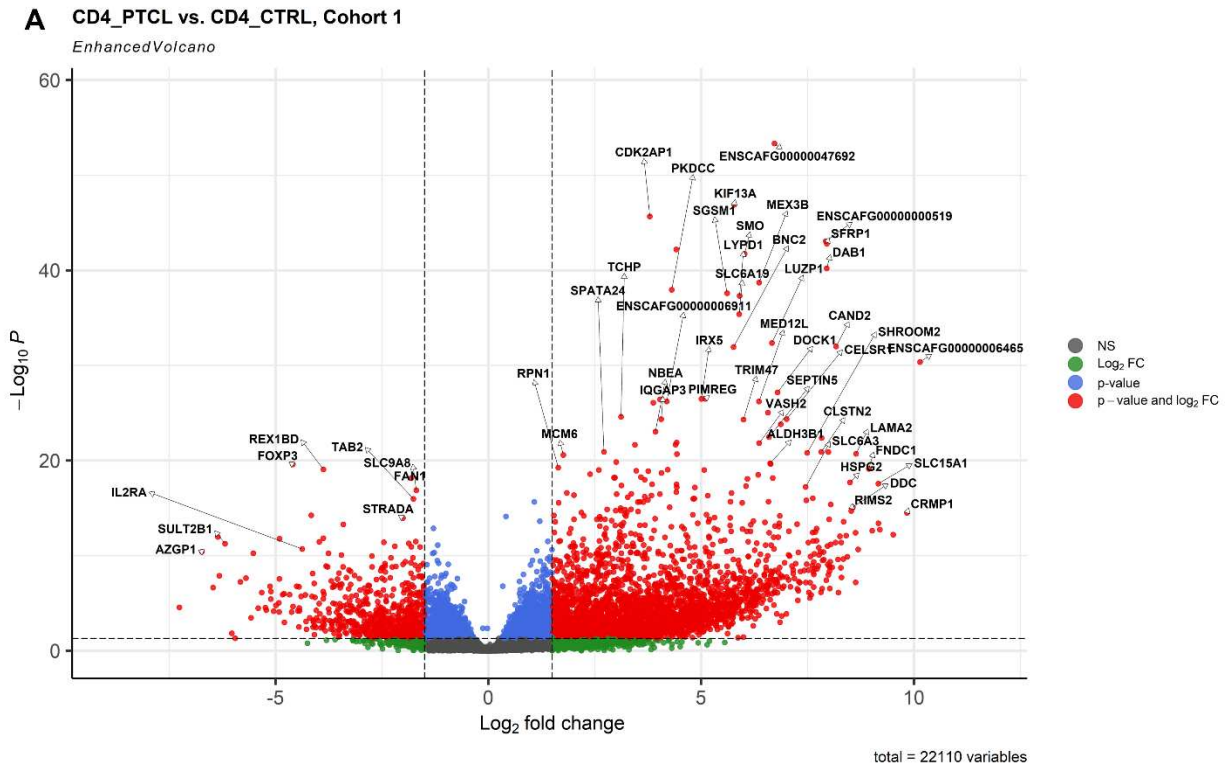
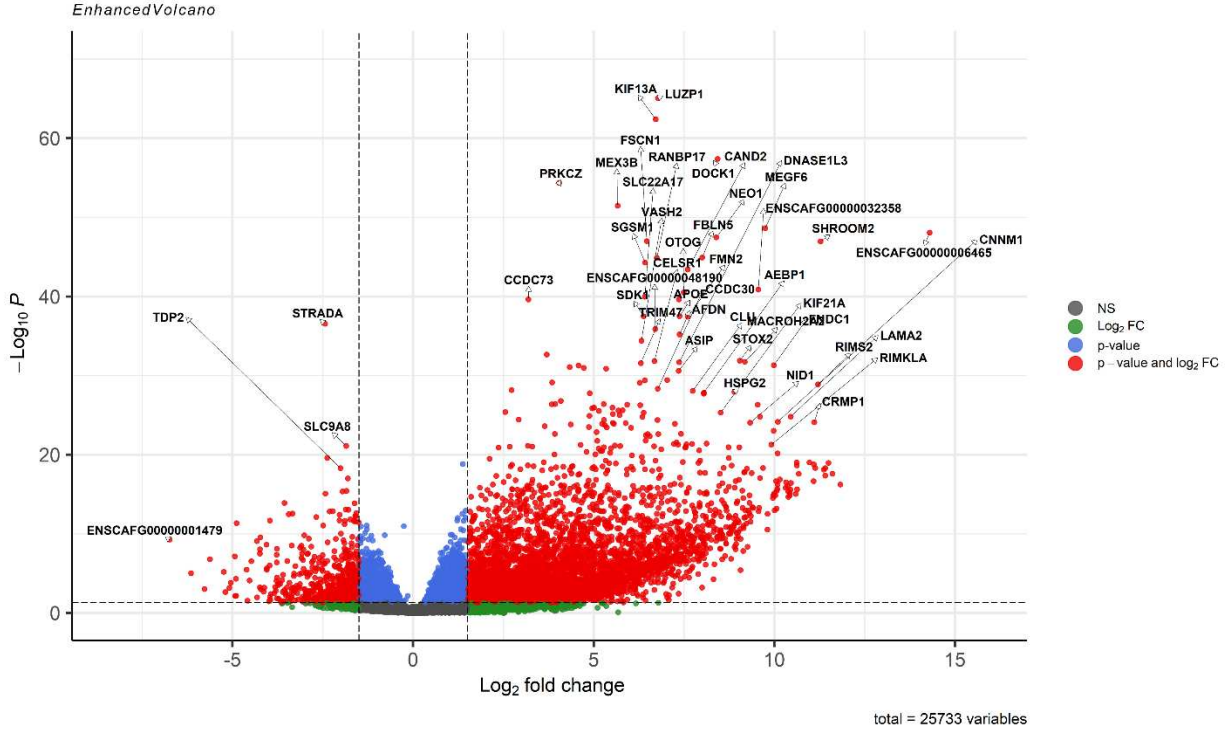


Figure 7: Unsupervised hierarchical clustering of variance stabilized transformed count data from CD4+ PTCL samples and control $CD4^+$ lymphocytes (**A**, Cohort 1) or control $CD4^+$ lymphocytes and thymocytes (**B**, Cohort 2) using average Euclidean distance and Ward clustering of the top 2000 genes with the highest median absolute derivation. Each row corresponds to a gene in the top 2000 by median absolute derivation, and each column corresponds to a dog in the cohort. Upregulated genes are shown in red, while downregulated genes are shown in blue. White represents no change in gene expression. All CD4+ PTCL samples demonstrate a gene expression profile that is distinct from that of the control CD4+ lymphocytes and thymocytes.

For both cohorts, genes were determined to be significantly differentially expressed based on a *p*-value cutoff of 0.05 following Benjamini–Hochberg adjustment. In Cohort 1, 4,807 genes and 1,977 genes were found to be significantly up- and downregulated, respectively, in canine $CD4^+$ PTCL compared to control $CD4^+$ lymphocytes (Figure 8A). In Cohort 2, the number of genes significantly up- and downregulated in canine $CD4^+$ PTCL compared to control $CD4^+$ lymphocytes was 6,651 and 2,195, respectively (Figure 8B), while 2,903 genes were significantly upregulated and 295 genes were significantly downregulated in $CD4^+$ PTCL compared to control $CD4^+$ thymocytes (Figure 8C).


A
CD4_PTCL vs. CD4_CTRL, Cohort 1
EnhancedVolcano
−Log10 P
Log2 fold change
total = 22110 variables
NS
Log2 FC
p-value
p − value and log2 FC
CDK2AP1
ENSCAFG00000047692
PKDCC
KIF13A
SGSM1
MEX3B
ENSCAFG00000000519
SMO
SFRP1
LYPD1
BNC2
DAB1
TCHP
SLC6A19
LUZP1
SPATA24
ENSCAFG00000006911
CAND2
MED12L
SHROOM2
IRX5
DOCK1
CELSR1
ENSCAFG00000006465
RPN1
NBEA
TRIM47
IQGAP3
PIMREG
SEPTIN5
VASH2
CLSTN2
MCM6
ALDH3B1
LAMA2
REX1BD
TAB2
SLC6A3
FOXP3
SLC9A8
FNDC1
HSPG2
SLC15A1
FAN1
DDC
IL2RA
RIMS2
CRMP1
STRADA
SULT2B1
AZGP1


B
CD4_PTCL vs. CD4_LN_CTRL, Cohort 2
EnhancedVolcano
−Log10 P
Log2 fold change
total = 25733 variables
NS
Log2 FC
p-value
p − value and log2 FC
KIF13A
LUZP1
FSCN1
RANBP17
CAND2
DNASE1L3
MEX3B
PRKCZ
DOCK1
MEGF6
SLC22A17
NEO1
VASH2
ENSCAFG00000032358
SGSM1
FBLN5
SHROOM2
CNNM1
OTOG
ENSCAFG00000006465
CELSR1
FMN2
AEBP1
CCDC73
ENSCAFG00000048190
SDK1
APOE
CCDC30
KIF21A
TDP2
STRADA
TRIM47
AFDN
MACROH2A2
PDC1
CLU
STOX2
LAMA2
ASIP
RIMS2
RIMKLA
NID1
HSPG2
CRMP1
SLC9A8
ENSCAFG00000001479

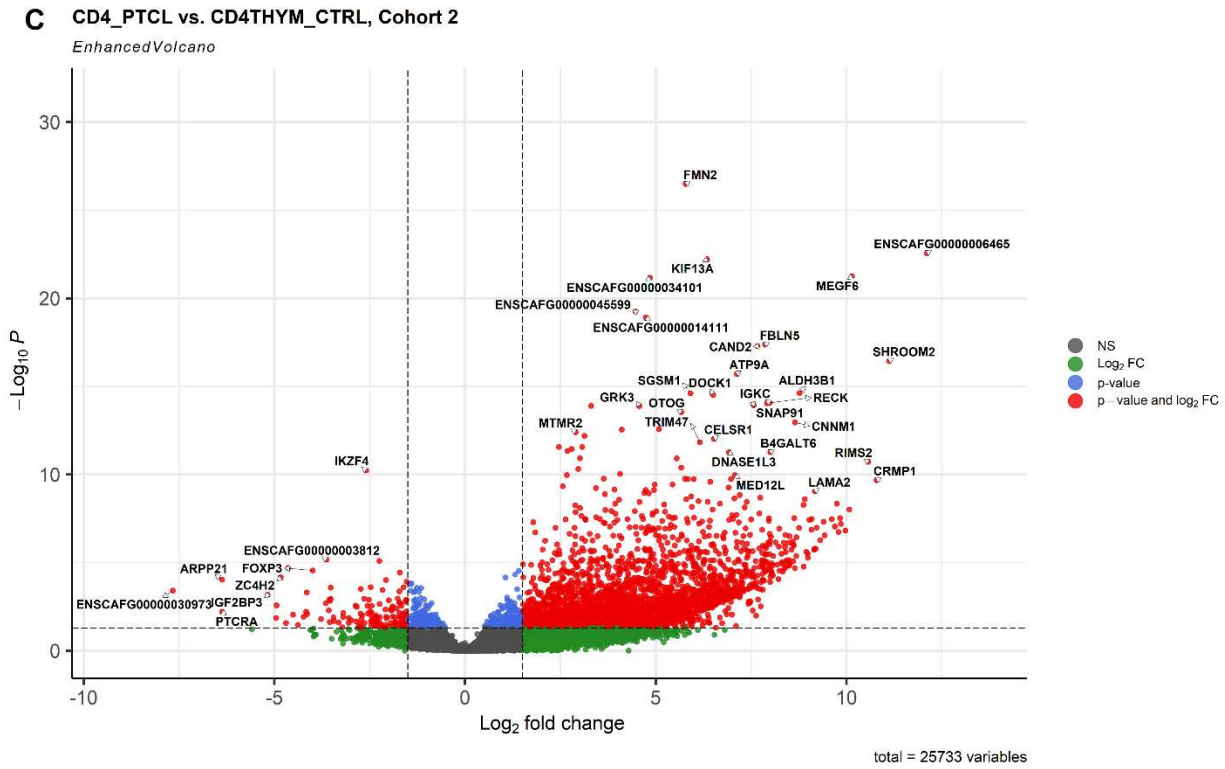


Figure 8: Volcano plot illustrating the top differentially expressed genes in canine CD4+ PTCL compared to control canine CD4+ lymphocytes (**A**, Cohort 1) or control CD4+ lymphocytes and thymocytes (**B**, Cohort 2). Upregulated and downregulated genes (defined as having a log2 fold change of > 1.5 and < -1.5, respectively) are delineated by the dotted vertical lines, and the red data points indicate those genes that were significantly upregulated or downregulated (based on a Benjamini–Hochberg adjusted p-value cutoff of 0.05). Blue data points indicate those genes that met our cutoff for significance but had a log2 fold change with an absolute value less than 1.5. Green data points indicate those genes that met our cutoff for log2 fold change but were not statistically significant. Gray data points indicate those genes that met neither our log2 fold change nor our p-value cutoff.

By GSEA, the top significantly enriched MSigDB gene sets (based on a *p*-value cutoff of 0.05 following Benjamini–Hochberg adjustment) in canine $CD4^+$ PTCL compared to control canine $CD4^+$ T-cells included various cancer gene signatures; receptor tyrosine kinase, MTOR, and KRAS signaling signatures; several cell cycle associated gene neighborhoods; and signatures associated with the extracellular matrix and angiogenesis (Figure 9).

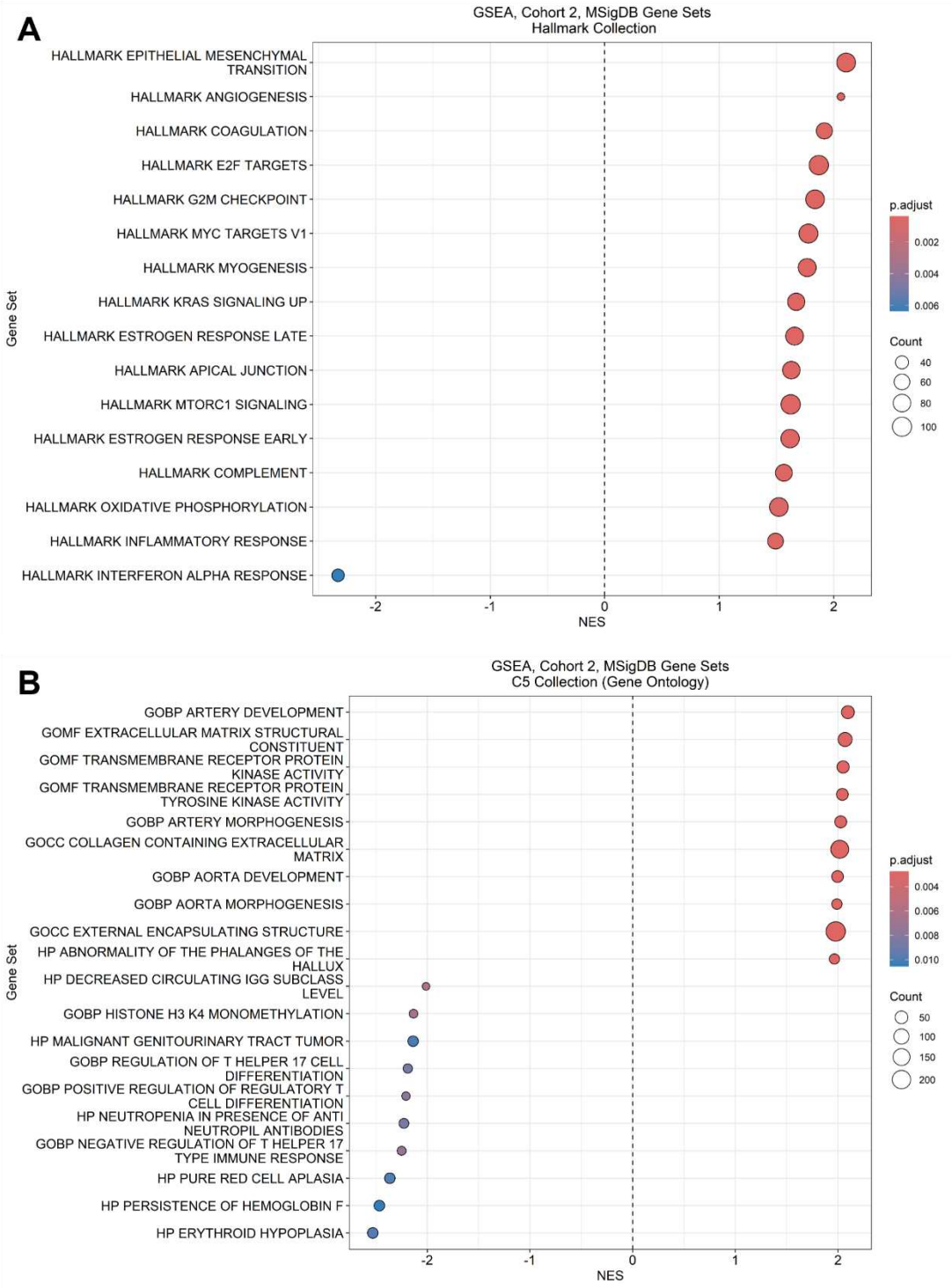


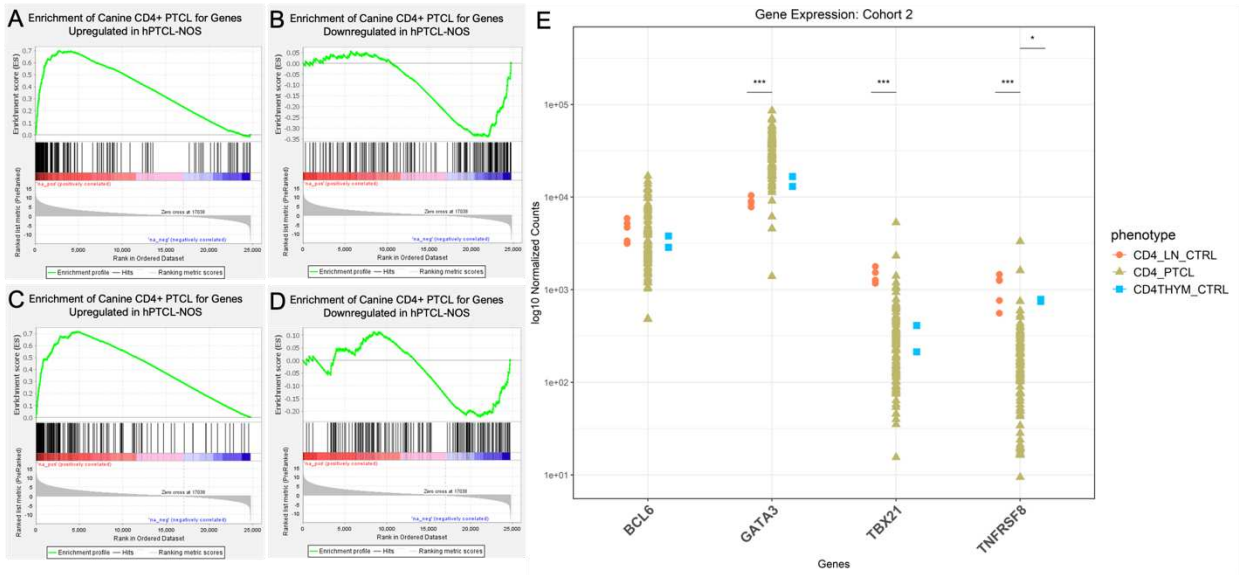
Figure 9: The top significantly enriched gene sets in the Hallmark (**A**) and Gene Ontology (**B**) collections of the Molecular Signatures Database in canine CD4⁺ PTCL compared to control CD4⁺ lymphocytes. NES = normalized enrichment score. Based on gene expression data from Cohort 2.

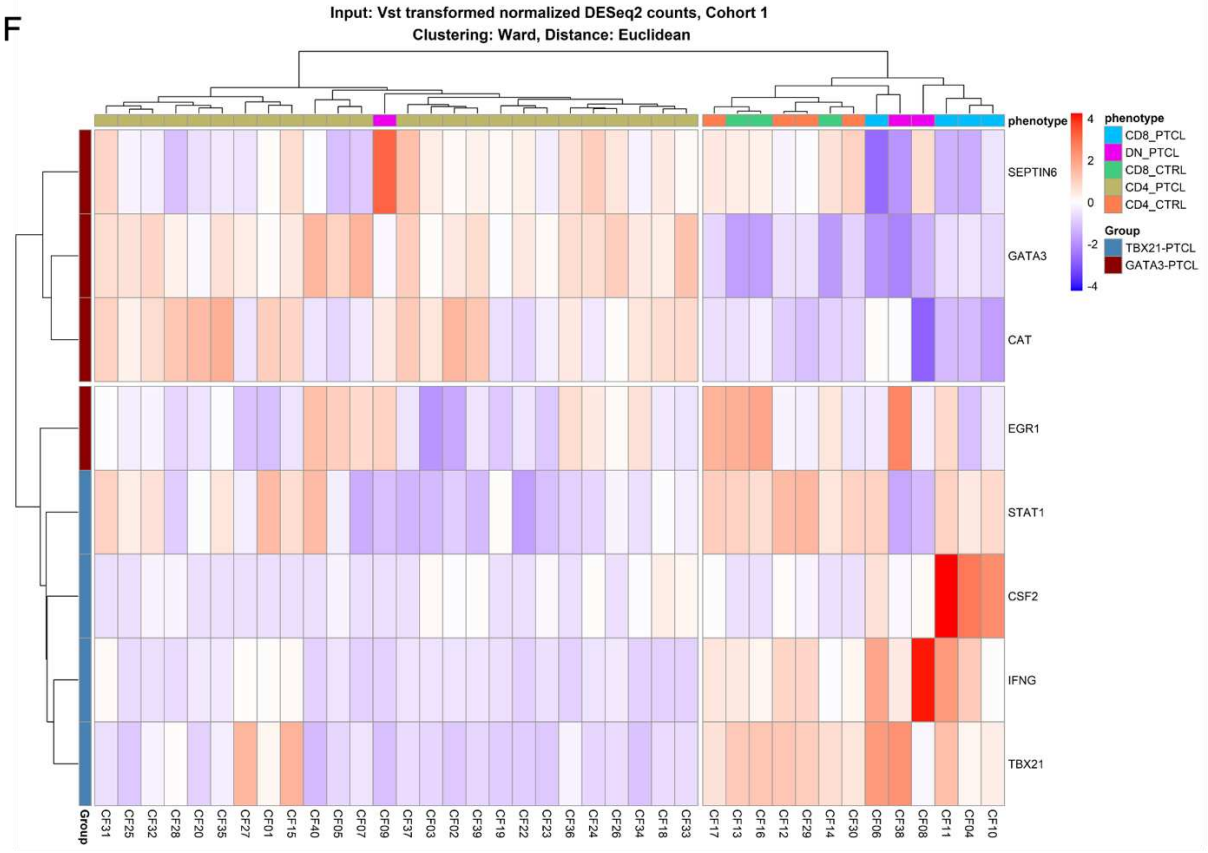
Canine CD4+ PTCL exhibits a molecular signature similar to the GATA3-PTCL subtype of PTCL-NOS in humans

GSEA using the ranked list of differentially expressed genes in canine CD4+ PTCL compared to control canine CD4+ lymphocytes revealed significant enrichment of canine CD4+ PTCL for the top 250 upregulated genes in human PTCL-NOS derived from Piccaluga et al.[31] and Etebari et al.[32], and negative enrichment for the 250 most downregulated genes in human PTCL-NOS, although in Cohort 1 the latter was only significant in the gene set derived from Piccaluga et al. (Table 2, Figure 10A-D). Canine CD4+ PTCL demonstrated increased expression of the genes associated with the GATA3-PTCL subtype and decreased expression of the genes associated with TBX21-PTCL subtype of human PTCL-NOS, as defined by Iqbal et al (Figure 10E). Compared to control canine CD4+ lymphocytes, canine CD4+ PTCL had significantly increased expression of GATA3 and decreased expression of TBX21 (Figure 10F).

Table 2: Enrichment scores for human PTCL-NOS gene signatures. Enrichment analysis was performed using the ranked list of differentially expressed genes identified between canine CD4+ PTCL cells and control canine nodal CD4+ lymphocytes.

Cohort 1			
Gene Set	**Normalized enrichment score**	***p*-value**	**FDR**
GSE6338: Top 250 upregulated genes in human PTCL-unspecified compared to control CD4+ T-cells	2.31	<0.001	<0.001
GSE6338: Top 250 downregulated genes in human PTCL-unspecified compared to control CD4+ T-cells	-1.98	<0.001	<0.001
GSE132550: Top 250 upregulated genes in human PTCL-NOS compared to control CD4+ T-cells	2.42	<0.001	<0.001
GSE132550: Top 250 downregulated genes in human PTCL-NOS compared to control CD4+ T-cells	-1.14	0.2	0.13
Cohort 2			
Gene Set	**Normalized enrichment score**	***p*-value**	**FDR**
GSE6338: Top 250 upregulated genes in human PTCL-unspecified compared to control CD4+ T-cells	2.18	<0.001	<0.001
GSE6338: Top 250 downregulated genes in human PTCL-unspecified compared to control CD4+ T-cells	-1.77	<0.001	<0.001
GSE132550: Top 250 upregulated genes in human PTCL-NOS compared to control CD4+ T-cells	2.23	<0.001	<0.001
GSE132550: Top 250 downregulated genes in human PTCL-NOS compared to control CD4+ T-cells	-1.01	<0.001	0.33


A Enrichment of Canine CD4+ PTCL for Genes Upregulated in hPTCL-NOS
B Enrichment of Canine CD4+ PTCL for Genes Downregulated in hPTCL-NOS
C Enrichment of Canine CD4+ PTCL for Genes Upregulated in hPTCL-NOS
D Enrichment of Canine CD4+ PTCL for Genes Downregulated in hPTCL-NOS
Enrichment score (ES)
Ranked list metric (PreRanked)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores
E
Gene Expression: Cohort 2
log10 Normalized Counts
Genes
BCL6
GATA3
TBX21
TNFRSF8
phenotype
CD4_LN_CTRL
CD4_PTCL
CD4THYM_CTRL


F
Input: Vst transformed normalized DESeq2 counts, Cohort 1
Clustering: Ward, Distance: Euclidean
phenotype
SEPTIN6
GATA3
CAT
EGR1
STAT1
CSF2
IFNG
TBX21
Group
phenotype
CD8_PTCL
DN_PTCL
CD8_CTRL
CD4_PTCL
CD4_CTRL
Group
TBX21-PTCL
GATA3-PTCL

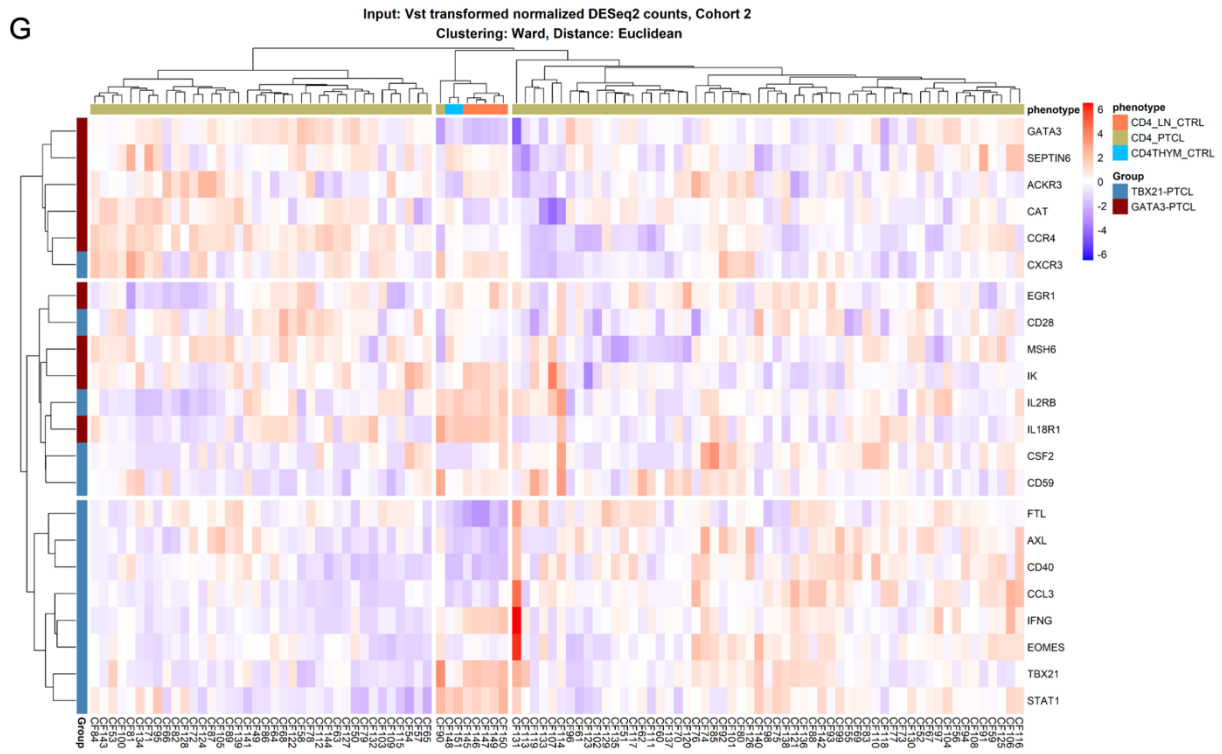


Figure 10: **The gene expression profile of canine CD4+ PTCL more closely resembles the GATA3-PTCL subgroup of human PTCL-NOS than the TBX21-PTCL subgroup.** Data from Cohort 2. (**A-D**) GSEA enrichment plots demonstrating the significant enrichment of the ranked list of differentially expressed genes in canine CD4+ PTCL for the top 250 upregulated genes in human PTCL-NOS (**A, C**) and negative enrichment for the 250 most downregulated genes in human PTCL-NOS (**B, D**), as obtained from GSE6338 and GSE132550. The ranking metric (gray) measures a gene's correlation with either the $CD4^+$ PTCL or control $CD4^+$ T cell phenotype. The vertical lines correspond to where members of the reference gene list appear in the ranked list of canine $CD4^+$ PTCL genes. If a gene in $CD4^+$ PTCL is also in the reference set, the enrichment score increases proportional to the ranking metric. The running enrichment score is represented by the green line. (**E**) A Wilcoxon signed-rank test comparing the means of normalized count data between $CD4^+$ PTCL and control nodal $CD4^+$ lymphocytes and $CD4^+$ thymocytes in Cohort 2 identified significantly increased expression of *GATA3*, significantly decreased expression of *TBX21* and ALCL marker *TNFRSF8*, and no difference in expression of AITL marker *BCL6*. These results are consistent with the results of DESeq2 differential expression analysis comparing canine $CD4^+$ PTCL to control nodal $CD4^+$ lymphocytes (*GATA3*: log2 fold change = 1.82, padj = 1.20×10^{-6}; *TBX21:* log2 fold change = -1.87, padj = 0.006; *TNFRSF8:* log2 fold change = -2.15 , padj = 0.002; and *BCL6:* log2 fold change = -0.003, padj = 0.99). (**F-G**) Hierarchical clustering of canine PTCL samples and control canine $CD4^+$ and $CD8^+$ lymphocytes based on expression levels of genes associated with the human GATA3-PTCL and TBX21-PTCL subtypes of PTCL-NOS (from Iqbal et al.). Hierarchical clustering was done using average Euclidean distance and Ward clustering of variance stabilized transformed normalized RNA-seq count data. Upregulated genes are shown in red, while downregulated genes are shown in blue. The majority of $CD4^+$ PTCL cases form Cohort 1 (**F**) show preferential expression of the genes associated with the GATA3-PTCL subtype, while other canine PTCL immunophenotypes are more heterogeneous. A subset of canine $CD4^+$ PTCL cases (left) from Cohort 2 (**G**) also tends to express genes associated with the GATA3-PTCL subtype, including *GATA3, SEPTIN6, ACKR3, CAT, CCR4,* and *MSH6*, while demonstrating relatively lower expression of TBX21-PTCL subtype-associated genes, including *IL2RB, CD59, CCL3, IFNG, EOMES, TBX21,* and *STAT1*. Interestingly, the subset of $CD4^+$ PTCL clustering together on the right of the heatmap exhibits more heterogeneous expression for both GATA3-PTCL and TBX21-PTCL gene signatures.

By GSEA, canine $CD4^+$ PTCL in both cohorts was also significantly enriched for gene signatures associated with hallmark PI3K/AKT (WP4172[46]) and MTORC1 signaling[30], upregulation of AKT (GSE1413[47]) and mTOR (GSE1413[47]), and downregulation of PTEN (GSE7562[48]) (Table 3,

Figure 11). Enrichment for PI3K/AKT/mTOR signaling and loss of PTEN are features described in the GATA3-PTCL subtype of human PTCL-NOS.[3,5]

Table 3: Enrichment scores for canine CD4+ PTCL compared to oncogenic signaling pathway gene signatures.

Cohort 1			
Gene Set	**Normalized enrichment score**	***p*-adj**	**FDR**
Hallmark PI3K/AKT signaling (WP4172)	1.42	0.01	0.009
MTORC1 Signaling	1.64	0.003	0.002
Downregulation of PTEN (GSE7562)	1.57	0.007	0.006
Upregulation of AKT (GSE1413)	1.48	0.04	0.03
Upregulation of mTOR (GSE1413)	1.53	0.01	0.01
Cohort 2			
Gene Set	**Normalized enrichment score**	***p*-adj**	**FDR**
Hallmark PI3K/AKT signaling (WP4172)	1.42	0.003	0.002
MTORC1 Signaling	1.62	0.002	0.001
Downregulation of PTEN (GSE7562)	1.64	0.002	0.001
Upregulation of AKT (GSE1413)	1.47	0.004	0.003
Upregulation of mTOR (GSE1413)	1.46	0.006	0.004

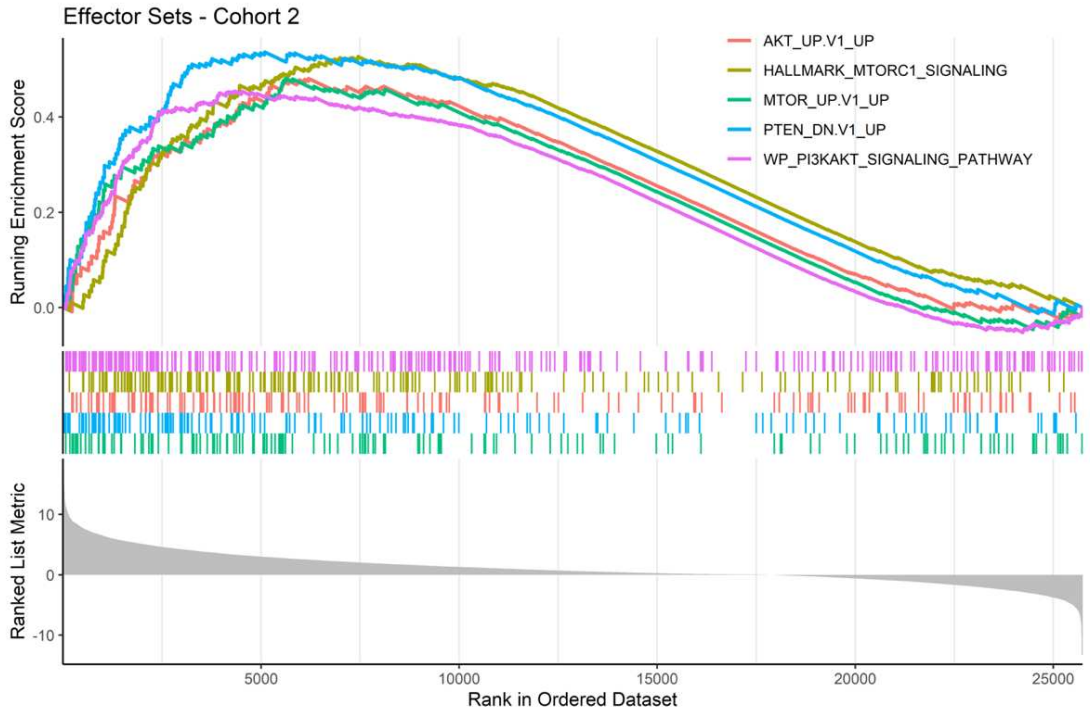


Figure 11: GSEA enrichment plot demonstrating the significant enrichment (with a Benjamini–Hochberg p-value cutoff of 0.05) of the gene expression profile of canine CD4+ PTCL for gene signatures associated with downregulation of PTEN and upregulation of the PI3K/AKT/mTOR signaling cascade. Red = Enrichment of canine CD4+ PTCL for the gene set upregulated with upregulation of AKT (GSE1413); yellow = Enrichment of canine CD4+ PTCL for the gene set

upregulated through activation of the mTORC1 complex (hallmark MTORC1 signaling pathway); green = Enrichment of canine CD4+ PTCL for the gene set upregulated with upregulation of MTOR (GSE1413); blue = Enrichment of canine CD4+ PTCL for the gene set upregulated with downregulation of PTEN (GSE7562); pink = Enrichment of canine CD4+ PTCL for PI3K-Akt signaling pathway genes (Wikipathways WP4172).

The mutational profile of canine CD4+ PTCL resembles human PTCL-NOS

After filtering with GATK pipeline tools and excluding variants annotated as intronic, splice site, splice region, and those predicted to have only low or modifier effects, a total of 196,147 and 13,636 variants were called in RNA-seq reads of $CD4^+$ PTCL samples and control samples of $CD4^+$ nodal lymphocytes and thymocytes, respectively, with a median of 2,046 variants per sample in $CD4^+$ PTCLs and 1,900 variants in control samples (Supplemental Figure 3). These totals represented the cumulative number of 71,913 different unique variants across canine $CD4^+$ PTCL samples, and 8,265 different unique variants across all control samples. 4,427 shared variants—defined in our study as variants spanning the same start position and end position on the same chromosome and having the same variant classification (nonsense, frameshift, etc.) and variant type (SNP, INDEL)—were called in both canine $CD4^+$ PTCL samples and control samples (Supplemental Figure 3). These shared variants were subsequently filtered out of the tumor cohort prior to downstream analysis to allow a more focused analysis of variants specific to $CD4^+$ PTCL samples.

Following exclusion of variants that were also called in control samples, a total of 103,896 variants were identified in canine $CD4^+$ PTCL samples, accounting for 67,486 unique variants and a median of 832 variants per dog (Figure 12A). Missense mutations were the most common variant classification, and single nucleotide polymorphisms and insertions represented the most common variant types (Figure 12B-C). Among single nucleotide variants, C>T transitions were the most common (Figure 12D). Cancer-associated genes that were most commonly mutated in canine $CD4^+$ PTCL included mutations in tumor suppressors like *KMT2D, FANCA*, and *NCOR1* in 54%, 46%, and 38% of tumors, respectively, and growth factor receptors like *EGFR*, *PDGFRB,* and *ERBB2* in 43%, 45%, and 40% of tumors, respectively (Figure 12E).

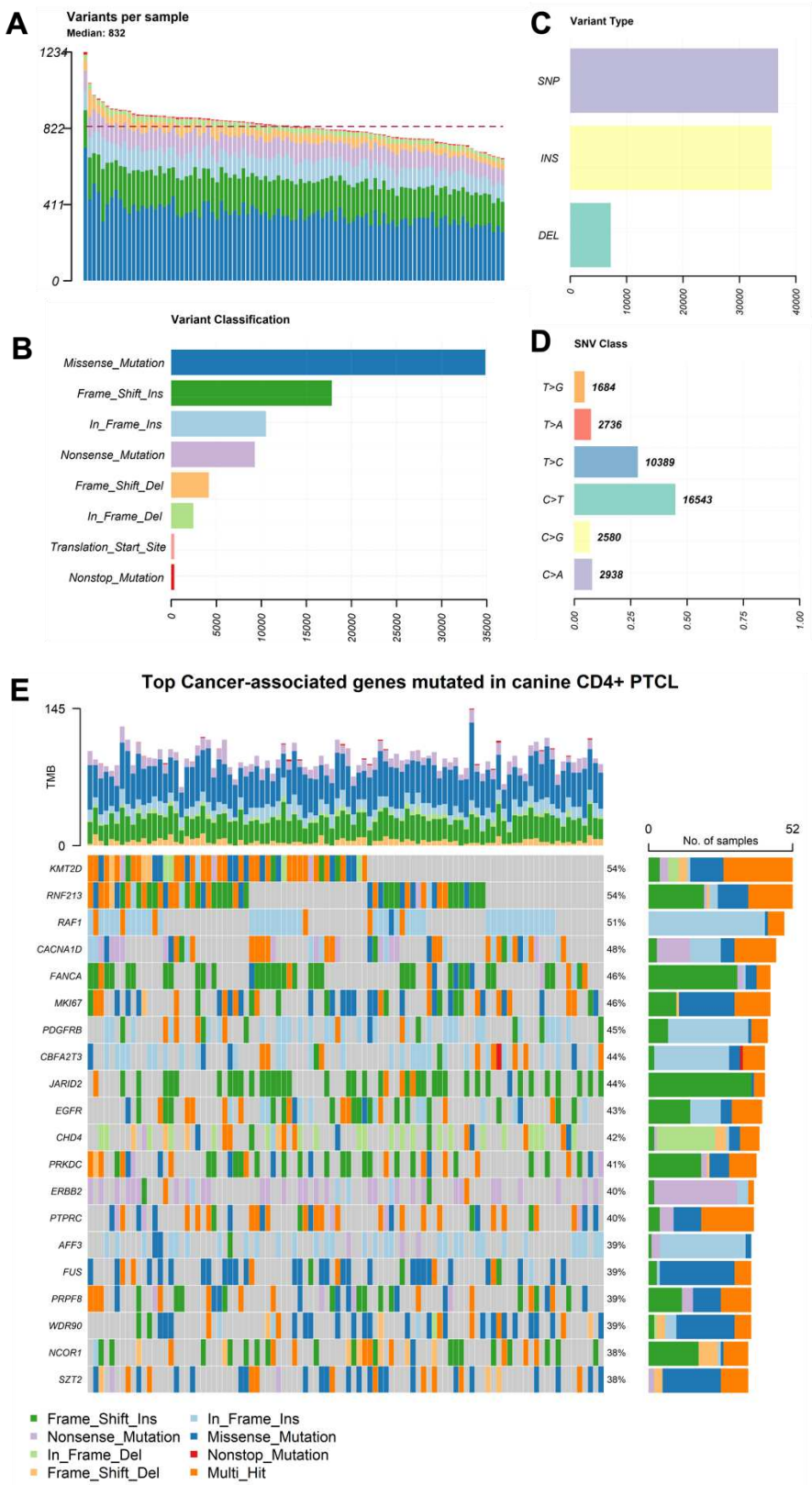
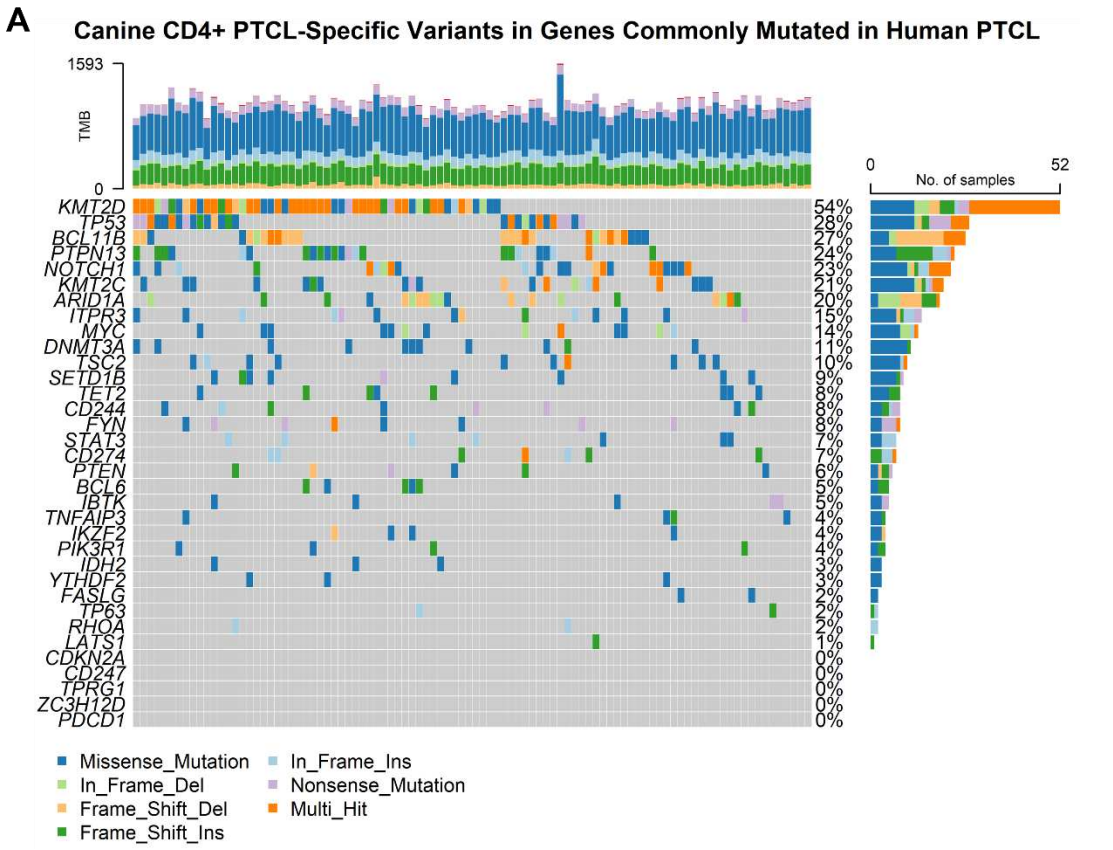


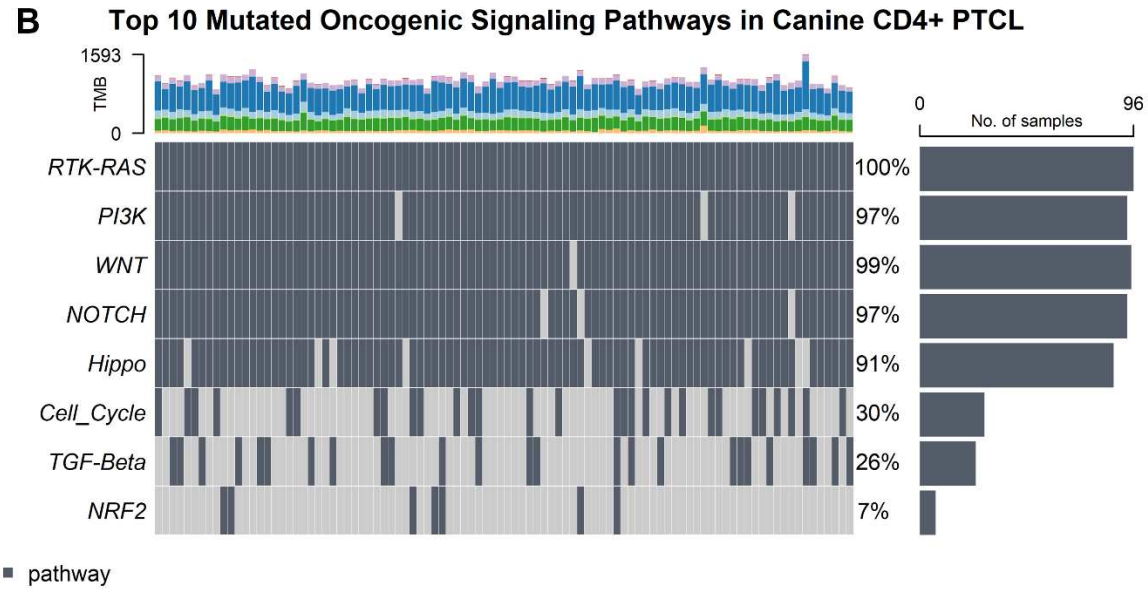
Figure 12: CD4⁺ PTCL-specific variants detected in bulk RNA-seq data from Cohort 2. The median number of variants per tumor sample was 832 (**A**), with missense mutations representing the most common classification (**B**). The

most common variant type was single nucleotide polymorphisms (SNPs) (**C**), and these were most often C>T transitions (**D**). (**E**) The top 20 genes from the OncoKB™ Cancer Gene List[41,42] mutated in CD4$^+$ PTCL included *KMT2D, RNF213, RAF1, CACNA1D, FANCA, MKI67, PDGFRB, CBFA2T3, JARID2, EGFR, CHD4, PRKDC, ERBB2, PTPRC, AFF3, FUS, PRPF8, WDR90, NCOR1,* and *SZT2*.

Among the most recurrently mutated genes in the GATA3-PTCL variant of PTCL-NOS are *TP53* and *PTEN*,[5,49,50] which were mutated in 28% and 6% of canine CD4$^+$ PTCL, respectively (Figure 13A). Gains in *MYC* and *STAT3* are also common in the GATA3-PTCL subtype, while gains in *BCL11B*, *BCL6, CD244, CD247, FASL6, TP63,* and *TPRG1* are common in TBX21-PTCL.[5] While several of these genes were mutated in a subset of canine CD4$^+$ PTCLs, RNA-seq cannot assess copy number aberrations. The presence of a *TP53* or *BCL11B* mutation in our study did not correlate with expression of other genes defining the GATA- and TBX21-PTCL subtypes (Supplemental Figure 4). Mutations affecting the PI3K signaling cascade, which is enriched in the human GATA3-PTCL subtype, were present in 97% of canine CD4$^+$ PTCL (Figure 13B).

Mutations in *DNTM3A* and *TET2* occur at comparable frequencies in the GATA3-PTCL and TBX21-PTCL variants,[5] and these were mutated in 11% and 8% of canine CD4$^+$ PTCL, respectively. *DNMT3A* in CD4$^+$ PTCL mutations tended to be distributed along the N-terminal region (4/11, 36%), the DNA methylase domain (2/11, 18%) or between domains (5/11, 45%), resembling the distribution of *DNMT3A* mutations in human PTCL-NOS (Figure 13C).[49] 14/32 (44%) of CD4$^+$ PTCL *TP53* mutations occurred in the DNA binding domain, which is also most commonly implicated in human PTCL-NOS (Figure 13D).[5] Other genes recurrently mutated in PTCL-NOS in general (and not specifically associated with either the GATA3- or TBX21-PTCL subtypes) were also mutated in canine CD4$^+$ PTCL. Most notably, a mutation in the histone methyltransferase *KMT2D* gene, which is mutated in 12% of human PTCL,[51] was present in 54% of canine CD4$^+$ PTCL (Figure 13A).


A
Canine CD4+ PTCL-Specific Variants in Genes Commonly Mutated in Human PTCL
1593
TMB
0
0
52
No. of samples
KMT2D 54%
TP53 28%
BCL11B 27%
PTPN13 24%
NOTCH1 23%
KMT2C 21%
ARID1A 20%
ITPR3 15%
MYC 14%
DNMT3A 11%
TSC2 10%
SETD1B 9%
TET2 8%
CD244 8%
FYN 8%
STAT3 7%
CD274 7%
PTEN 6%
BCL6 5%
IBTK 5%
TNFAIP3 4%
IKZF2 4%
PIK3R1 4%
IDH2 3%
YTHDF2 3%
FASLG 2%
TP63 2%
RHOA 2%
LATS1 1%
CDKN2A 0%
CD247 0%
TPRG1 0%
ZC3H12D 0%
PDCD1 0%
Missense_Mutation
In_Frame_Del
Frame_Shift_Del
Frame_Shift_Ins
In_Frame_Ins
Nonsense_Mutation
Multi_Hit


B
Top 10 Mutated Oncogenic Signaling Pathways in Canine CD4+ PTCL
1593
TMB
0
0
96
No. of samples
RTK-RAS 100%
PI3K 97%
WNT 99%
NOTCH 97%
Hippo 91%
Cell_Cycle 30%
TGF-Beta 26%
NRF2 7%
pathway

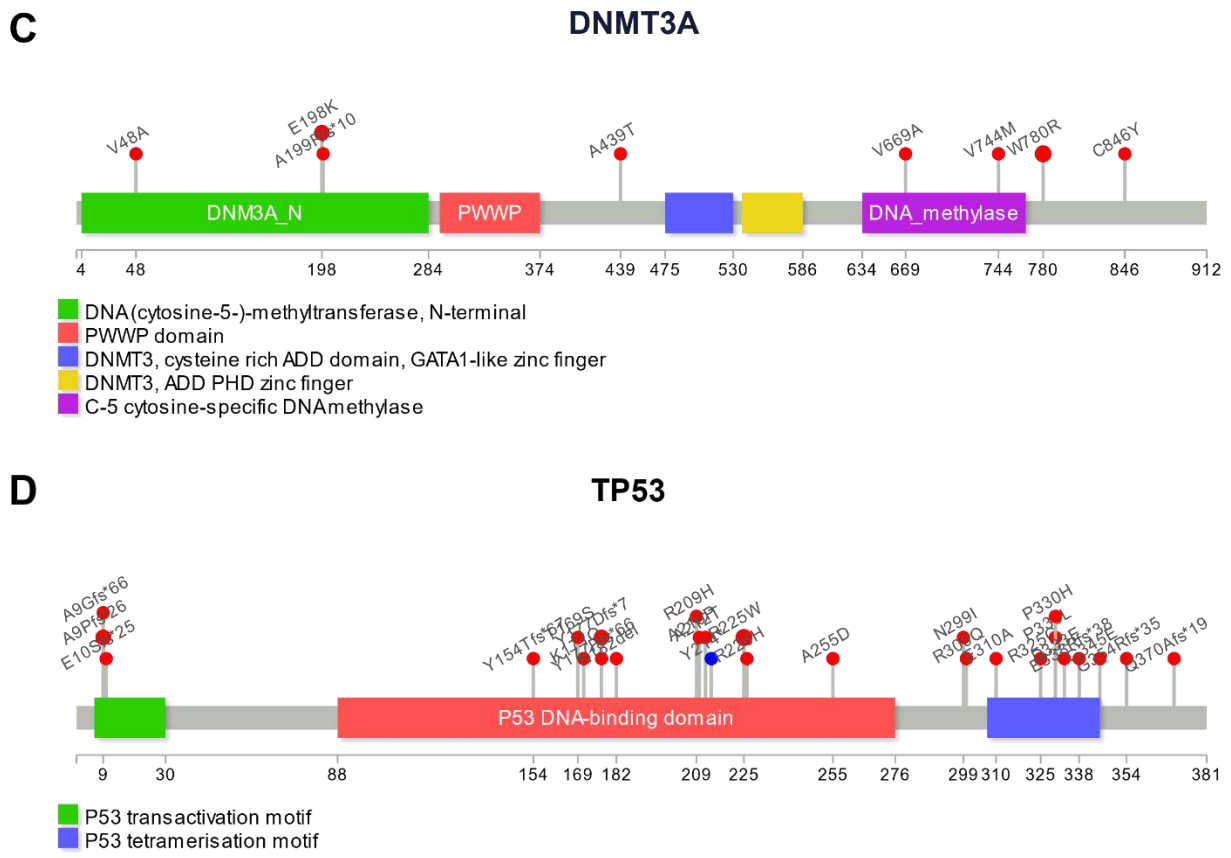


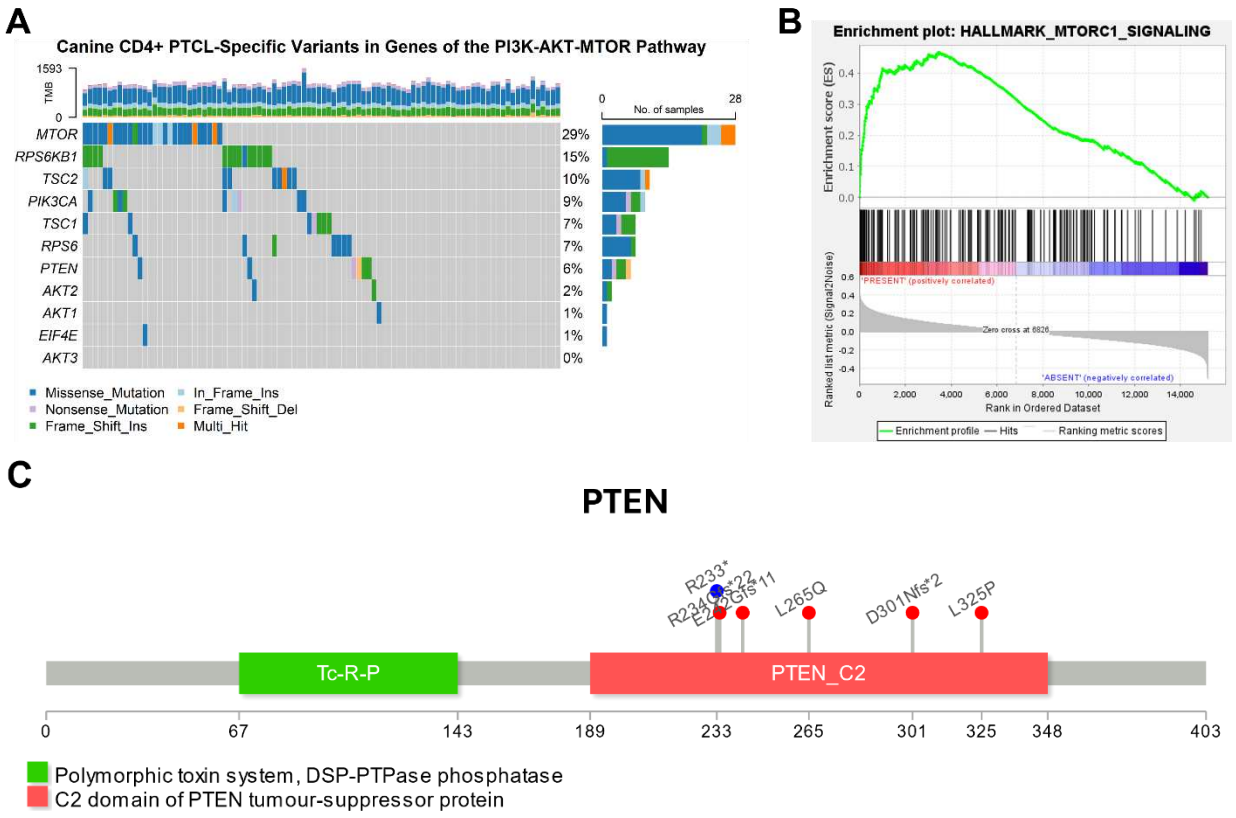
Figure 13: The mutational burden of canine CD4⁺ PTCL resembles human PTCL-NOS. Recurrently mutated genes in human PTCL that were also mutated in a subset of canine CD4⁺ PTCL included *TP53, DNMT3A, BCL11B, FYN, STAT3, CD244, TET2, BCL6, IBTK, MYC, TNFAIP3, PTEN, FASLG, TP63,* and *LATS1*. Mutations in *TP53, STAT3*, *MYC,* and *PTEN* are particularly associated with the GATA3-PTCL subtype, and these were mutated in 28%, 7%, 4%, and 3% of canine CD4⁺ PTCL, respectively (**A**). The PI3K signaling pathway is recurrently mutated in human PTCL-NOS, and this pathway was similarly implicated in canine CD4⁺ PTCL (**B**). (**C-D**) Annotated protein sequences for canine DNMT3A and canine TP53 were extracted from UniProt and used to evaluate the distribution of variant calls in these genes. While mutations in the DNA methylase domain of DNMT3A are reported for both variants of human PTCL-NOS, they are more commonly a feature of TBX21-PTCL, while GATA3-PTCL mutations tend to distribute along the N-terminal and PWWP domains. In canine CD4⁺ PTCL, 4/11 (36%) of DNM3A mutations occurred in the N-terminal region, 2/11 (18%) occurred in the DNA methylase domain, and 5/11 (45%) occurred between domains (**C**). In human PTCL-NOS, mutations in TP53 typically occur in the DNA binding domain, and 14/32 (44%) of canine CD4⁺ PTCL TP53 variants were also distributed along this domain (**D**).

The mutational profile of canine CD4⁺ PTCL informs enriched *MTOR* signaling

Given the enrichment of canine CD4⁺ PTCL and human PTCL-NOS for PI3K/AKT/mTOR signaling, we evaluated common genes along this pathway for mutations in canine CD4⁺ PTCL. *MTOR* was the most commonly mutated gene in this pathway, with variants in 29% of canine CD4⁺ PTCL (Figure 14A). Inactivating *PTEN* deletions are an important feature of human PTCL-NOS speculated to contribute to PI3K pathway enrichment in those tumors, but *PTEN* mutations were

less frequent seen in our study (6%) (Figure 14A), despite CD4$^+$ PTCLs being enriched for gene signatures associated with downregulation of PTEN activity (Figure 11). Importantly, RNA-seq data is limited in its ability to detect large deletions, so we may be underestimating the true prevalence of *PTEN* mutations in our canine tumors.

Mutations in *MTOR* resulted in greater enrichment for *MTOR* signaling gene signatures (Figure 14B), suggesting that these may be activating mutations, but no differences in enrichment for PI3K/AKT/mTOR signaling were observed between samples with or without *PTEN* mutations. *MTOR* mutations in CD4$^+$ PTCL tended to aggregate along the catalytic phosphatidylinositol 3- and 4-kinase and FAT domains (Figure 14C). Interestingly, all *PTEN* mutations in CD4$^+$ PTCL were located along the C2 domain (Figure 14D), which is important for membrane binding and activation of *PTEN*.[52] However, these mutations did not correlate with enrichment for gene signatures associated with downregulation of PTEN, suggesting that they may not impede PTEN activity.



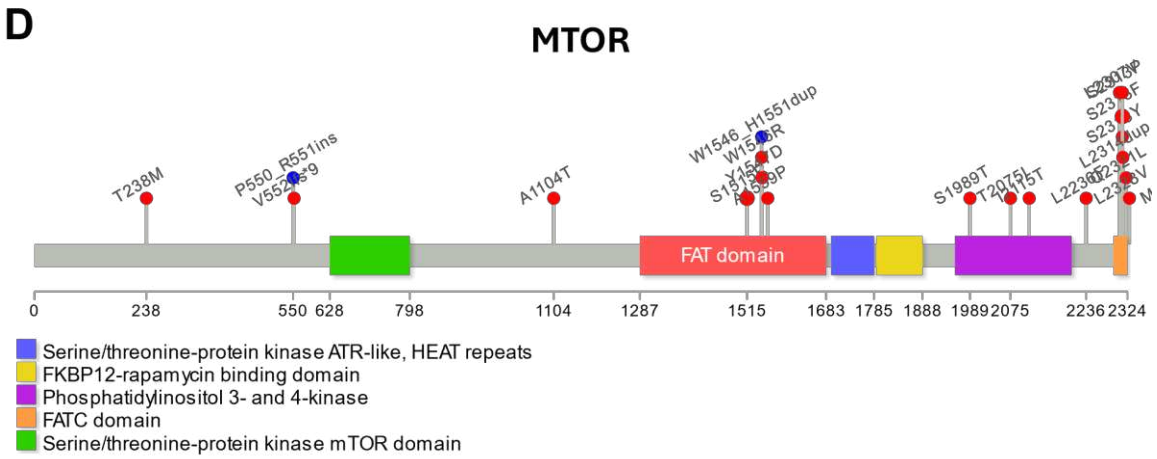


Figure 14: (**A**) Mutated genes along the PI3K-AKT-mTOR pathway in canine CD4$^+$ PTCL include *MTOR, RPS6KB1, TSC2, PIK3CA, TSC1, RPS6, PTEN, AKT2, AKT1,* and *EIF4E*. (**B**) CD4$^+$ PTCL with an *MTOR* mutation were significantly more enriched (NES = 1.74, p = 0.02) for genes associated with MTORC1 signaling, which could indicate an activating effect for these mutations. (**C**) All *PTEN* mutations in CD4$^+$ PTCL were located along the C2 domain important for membrane binding and activation of PTEN. (**D**) *MTOR* mutations in CD4$^+$ PTCL tended to aggregate along the catalytic phosphatidylinositol 3- and 4-kinase and FAT domains.

GATA3 upregulation in canine CD4$^+$ PTCL is not driven by *GATA3* mutations or mutations in upstream activators

Evaluation of upstream activators of *GATA3* revealed mutations in *SATB1*, *IL4*, *MTOR*, histone acetyltransferase *EP300*, β-catenin encoding *CTNNB1*, and *STAT6* in 41%, 34%, 29%, 17%, 9%, and 7% of CD4$^+$ PTCL, respectively (Figure 15A). At least one mutation in an upstream GATA3 activator was identified in 76/96 (79%) of CD4$^+$ PTCL. *GATA3* itself was mutated in 12% of CD4$^+$ PTCLs. The presence of a mutation in at least one upstream GATA3 activator did not result in significantly altered expression of *GATA3* transcript in CD4$^+$ PTCL (Figure 15B).

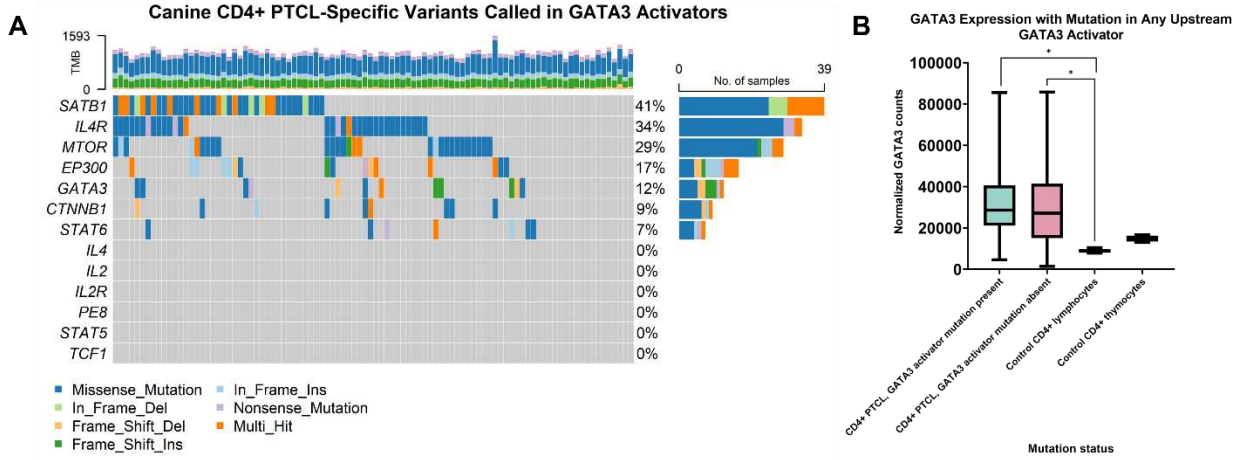


Figure 15: Mutations in GATA3 activators in CD4⁺ PTCL. (**A**) Mutations in several known upstream activators of GATA3 were mutated in canine CD4⁺ PTCL, including *SATB1, IL4R, MTOR, EP300, CTNNB1,* and *STAT6.* (**B**) The presence of a mutation in at least one upstream GATA3 activator did not significantly correlate with *GATA3* transcript expression in CD4⁺ PTCL based on a Tukey's multiple comparisons test on DESeq2 normalized counts of *GATA3* in canine CD4⁺ PTCL, control nodal CD4⁺ lymphocytes, and CD4⁺ thymocytes from Cohort 2.

Variant analysis of 96 canine CD4+ PTCLs corroborates previously reported canine T-cell lymphoma gene mutations

SATB1 and *PTEN* mutations have been previously reported to occur in up to 25% of canine T-cell lymphoma.[14] *SATB1* was mutated in 41% of our canine CD4⁺ PTCLs, although only 6% demonstrated mutations in *PTEN* (Figure 16A). *SATB1* mutations were most frequently dispersed along the CUT and CUT1-like DNA-binding domains (Figure 16B). In one study,[13] over 40% of canine T-cell lymphomas had mutations in *MET, KDR, STK11,* and *BRAF,* and these genes were mutated in 11%, 17%, 14%, and 4% of CD4⁺ PTCL in our study, respectively. The most frequently mutated gene in our study that had been previously reported in canine T-cell lymphoma was *EGFR*, which was found in 43% of our CD4⁺ PTCLs. Interestingly, this mutation was only found in 1/16 (6.25%) T-cell lymphomas in McDonald et al.[13] *EGFR* mutations in our study most often affected the Furin-like cysteine rich region, receptor L domain, and growth factor receptor domain IV (Figure 16C). No mutations were identified in *EEF1A1, KCND2,* and *PSMA1*, which were reported as significantly mutated genes in "Boxer type" T-cell lymphoma based on supplementary data from Elvers et al.[14]

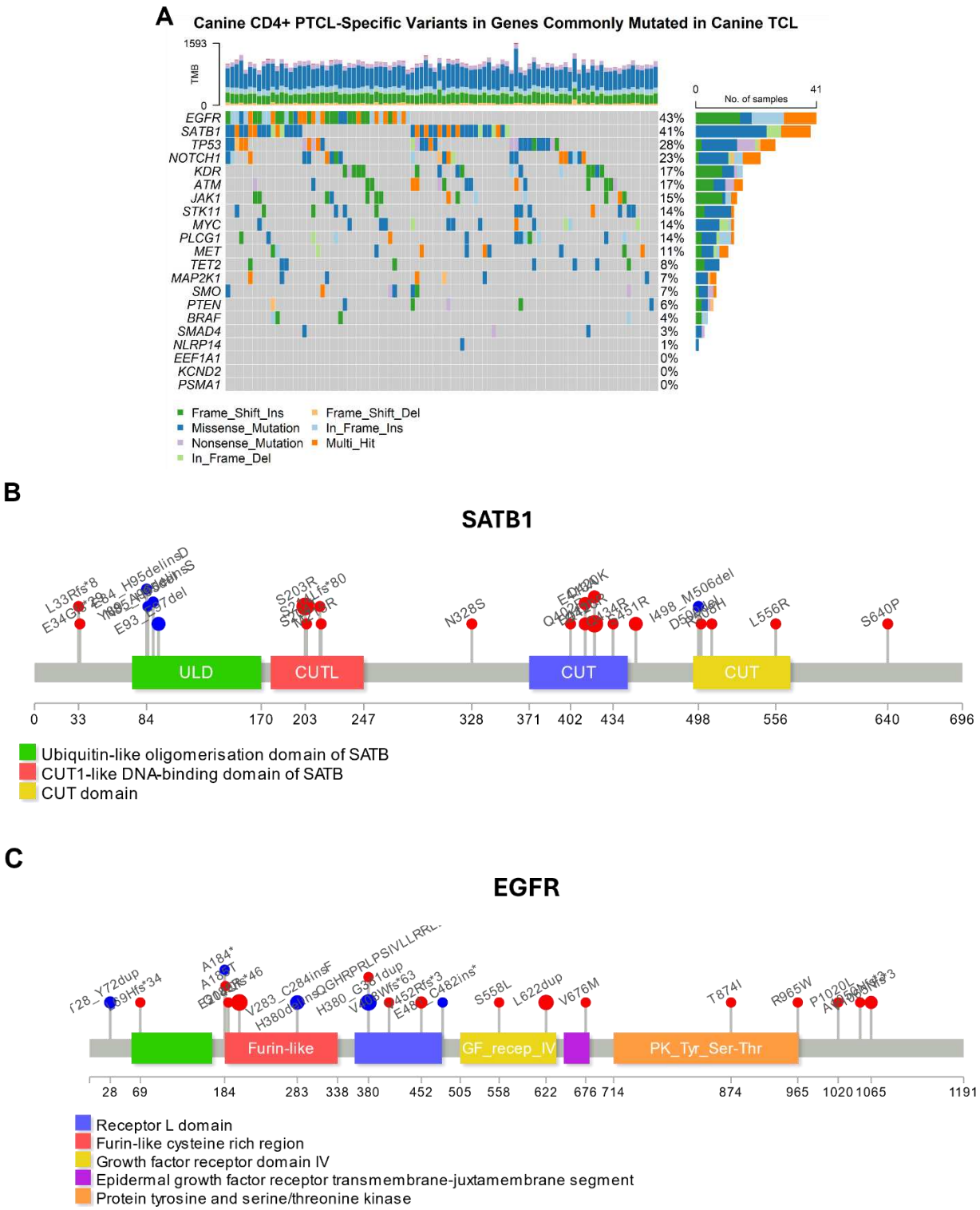


Figure 16: (**A**) Oncoplot of mutations in CD4⁺ PTCL cases from our study in genes reported in previous studies to be commonly mutated in canine T-cell lymphoma. (**B**) *SATB1* mutations, previously reported in up to 25% of canine T-cell lymphomas and present in 41% of canine CD4⁺ PTCLs in our study, tended to disperse along the CUT and CUT1-like DNA-binding domains, with fewer binding the ubiquitin-like oligomerization domain or binding regions between domains. (**C**) *EGFR* mutations, previously reported in 6% of canine T-cell lymphomas and present in 43% of canine CD4⁺ PTCLs in

our study, tended to disperse along the Furin-like cysteine rich region, receptor L domain, and growth factor receptor domain IV.

Discussion

In this study, we investigated the gene expression and mutational profile of canine PTCL, with a particular focus on whether canine PTCL molecularly resembles human PTCL-NOS. We found that the gene expression profile of canine $CD4^+$ PTCL resembles the human GATA3 PTCL-NOS subtype and in both species, these PTCL are enriched for the PI3K/AKT/MTOR pathway. Additionally, genes recurrently mutated in human PTCL-NOS, and particularly the GATA3-PTCL subtype, are also found in canine $CD4^+$ PTCL and affect similar protein-coding domains.

The immunophenotype most commonly diagnosed by flow cytometry in dogs, $CD4^+$ PTCL, exhibited the most consistent gene expression profile, whereas the gene expression profiles of $CD8^+$ and CD4-CD8- canine PTCLs suggested a different cell of origin. This observed heterogeneity may be due in part to the small sample sizes representing these phenotypes. We elected to focus on cases of the distinct $CD4^+$ PTCL immunophenotype for our study.

Our study revealed several similarities between canine PTCL and human PTCL-NOS, including similarities in clinical presentation, immunophenotype, gene expression, and mutated genes. In humans, a nodal presentation is most common with PTCL-NOS,[2] and peripheral lymphadenopathy was the most frequently reported clinical presentation of canine PTCL in our study (93%). In humans, there is a reported male predominance of 1.5-1.9:1,[2] and male dogs were similarly overrepresented in our study (65%). Hypercalcemia is a commonly observed feature on biochemistry profiles in human PTCL-NOS,[2] and 48% of canine PTCL cases in our study with biochemical data were hypercalcemic. One difference between human and canine PTCL is frequent mediastinal involvement of the canine tumor.

By flow cytometry, canine CD4$^+$ PTCLs expressed the pan-leukocyte marker CD45 and T-cell marker CD3, but loss of CD5 was observed in 42% of cases. Similarly, the majority of PTCL-NOS cases in humans lose expression of one or more mature T-cell antigens, most often CD7 or CD5.[2]

GSEA revealed significant enrichment of canine CD4$^+$ PTCL for the top 250 upregulated genes in human PTCL-NOS and significant negative enrichment for the top 250 downregulated genes in human PTCL-NOS based on a previous microarray study,[31] suggesting that canine CD4$^+$ PTCL and human PTCL-NOS are comparable diseases at a broad gene expression level.

Of the two recently described molecular subgroups of human PTCL-NOS, TBX21-PTCL and GATA3-PTCL, the gene expression profile of canine CD4$^+$ PTCL most closely resembled that of the GATA3-PTCL subtype. In humans, this subtype is associated with a poorer prognosis and characterized by increased expression of transcription factor GATA3 and its target genes.[3] Similarly, canine CD4$^+$ PTCL exhibited increased expression of *GATA3* and decreased expression of *TBX21*. In addition to being upregulated in the transcriptome, immunohistochemistry performed by our collaborator Lauren Harris identified increased expression of GATA3 protein in the peripheral nodes of canine CD4$^+$ PTCL samples compared to control lymph nodes from healthy dogs.[53] Canine CD4$^+$ PTCL also tended to upregulate other genes in the GATA3-PTCL signature (such as *IL18R1*, *CCR4*, *CAT*, and *ACKR3*) and downregulate other genes in the TBX21-PTCL signature (such as *IFNG*, *IL2RB*, *EOMES*, and *CCL3*) compared to other canine PTCL phenotypes and control canine CD4$^+$ and CD8$^+$ lymphocytes. Mutations in *GATA3* or at least one upstream *GATA3* activator were identified in 12% and 76% of CD4$^+$ PTCL, respectively. However, these mutations did not correlate with *GATA3* transcript expression, suggesting that *GATA3* overexpression in CD4$^+$ PTCL is driven by another mechanism.

In humans, the GATA3-PTCL subtype is also characterized by enrichment for the MYC and PI3K/AKT/mTOR signaling pathways that promote cellular proliferation,[3] and frequent mutations in the tumor suppressor gene *PTEN*.[5] Mutations in PI3K pathway-associated genes were identified in 97% of CD4$^+$ PTCL in our study, and GSEA revealed significant enrichment of the gene expression profile of canine CD4$^+$ PTCL for gene signatures associated with upregulation of the PI3K/AKT/mTOR pathway and downregulation of PTEN. This corroborates previous studies that found PTEN-mTOR pathway mutations in almost half of "Boxer type" canine T-cell lymphomas,[14] which are most likely the CD4$^+$ PTCL group being examined here. While *PTEN* mutations were only identified in 6% of CD4$^+$ PTCL in our study, variant calling using RNA-seq data is limited, and this may not represent the true prevalence of *PTEN* mutations in our cohort. Importantly, RNA-seq data cannot identify structural variants like large deletions.

Immunohistochemistry performed by our collaborator Lauren Harris identified diffuse expression of phosphorylated AKT in the lymph nodes of canine PTCL samples compared to control lymph nodes from healthy dogs.[53] Her work also demonstrated that inhibition of PI3K *in vitro* resulted in decreased proliferation and increased cell death in primary CD4$^+$ PTCL cells harvested from patient lymph nodes,[53] further illustrating the reliance of canine CD4$^+$ PTCL on PI3K/AKT/mTOR signaling for tumor cell proliferation and survival. These findings further support a similar molecular pathogenesis between the GATA3-PTCL subtype of human PTCL-NOS and canine CD4$^+$ PTCL. Given the current heterogeneity of clinical and molecular features in human PTCL-NOS, this narrowing down of the most specific subgroup resembled by canine CD4$^+$ PTCL should allow for more precise preclinical applications of our proposed canine model.

While we have demonstrated a similar dependence on PI3K/AKT signaling between human GATA3-PTCL and canine CD4$^+$ PTCL, future studies are warranted to clarify the mechanism and specific downstream effectors implicated in this dependence. The PI3K/AKT/mTOR axis is known

to promote cell survival and proliferation through a variety of mechanisms, including through the inhibition of pro-apoptotic proteins like BAD and FOXO[54,55] and increased synthesis of cyclins,[56,57] among others. Notably, *FOXO1* was downregulated in our neoplastic CD4$^+$ PTCL cells compared to control nodal CD4$^+$ T cells (log2 fold change = -1.2, padj = 0.01 for Cohort 1 and log2 fold change = -1.2, padj = 0.005 for Cohort 2). PI3K/AKT signaling also promotes the activity of glycolytic enzymes[58,59] and the upregulation of glucose transporters,[60,61] which helps produce the cellular metabolic shift toward aerobic glycolysis observed in many cancer cells.[62] One such glucose transporter influenced by PI3K/AKT signaling is GLUT4,[60] encoded by the *SLC2A4* gene. In our study, this gene was upregulated in canine CD4$^+$ PTCL cells compared to control nodal CD4$^+$ T cells (log2 fold change = 2.9, padj = 0.01 for Cohort 1, and log2 fold change = 2.7, padj = 0.08 for Cohort 2). For these reasons, we suspect the dependency of CD4$^+$ PTCL cells on this signaling cascade is multifactorial, although additional studies are needed. In human PTCL-NOS, much of the effort to understand the therapeutic applications of the PI3K/AKT/mTOR dependency has been focused on more upstream targets, like utilizing PI3K inhibitors,[63] and to our knowledge investigations into more specific downstream effectors remain limited. The correlation between GATA3 upregulation and enrichment for PI3K/AKT/mTOR signaling both in human GATA3-PTCL and canine CD4$^+$ PTCL could reflect a molecular connection between these factors, a finding suggested by a previous study noting that GATA3 target genes in both immature and mature T-cell neoplasms consistently included genes involved in PI3K/AKT signaling.[64] Additionally, as previously mentioned, deletions involving the PI3K inhibitor *PTEN* are more prevalent in the GATA3-PTCL subtype of human PTCL-NOS than the TBX21-PTCL subtype,[5] suggesting that the PI3K/AKT/mTOR enrichment frequently seen in tumors of this subtype may result from an increased susceptibility of this subtype to deletions in *PTEN*. This correlation seems similarly probable in canine CD4$^+$ PTCL based on whole exome sequencing data identifying PTEN-mTOR pathway mutations in almost half of “Boxer type” canine T-cell lymphomas,[14] and the frequent mutations in the PI3K signaling cascade identified by

our own variant analysis. However, the relationship (if any) between GATA3 upregulation and this increased susceptibility to *PTEN* deletions remains to be explored.

Variants identified in 96 of the CD4$^+$ PTCLs in our study corroborated previously reported recurrent mutations in canine T-cell lymphoma and identified overlap between genes commonly mutated in human PTCL-NOS and canine CD4$^+$ PTCL, particularly those associated with the GATA3-PTCL subtype. In addition to the *PTEN* mutations discussed above, recurrent mutations in *SATB1* have been previously reported in up to 25% of "Boxer type" canine T-cell lymphomas. We identified *SATB1* mutations in 41% of CD4$^+$ PTCLs in our cohort. Interestingly, 43% of CD4$^+$ PTCLs in our study had a mutation in *EGFR*, which was previously identified in only one case of canine T-cell lymphoma.[13]

Among the most recurrently mutated genes in human PTCL-NOS are *TP53*, *DNMT3A, STAT3, TET2, MYC,* and *PTEN*,[5,49,50,65,66] which were mutated in 28%, 10%, 7%, 5%, 4%, and 3% of canine CD4$^+$ PTCLs, respectively. While *DNMT3A* mutations are fairly ubiquitous among human GATA3-PTCL, TBX21-PTCL, and AITL, mutations in the *CDKN2A/TP53* axis, loss-of-function mutations and deletions in *PTEN*, and gains or amplifications of *STAT3* and *MYC* are particularly characteristic of GATA3-PTCL. The GATA3-PTCL subtype of human PTCL-NOS is also particularly prone to mutation in genes along the PI3K signaling pathway, which was similarly implicated in 81% of canine CD4$^+$ PTCLs in our study. As mentioned previously, RNA-seq data cannot be used to detect large deletions, which likely underestimates the true frequency of *PTEN* mutations in CD4$^+$ PTCL.

In human TBX21-PTCL, mutations in the DNA methylase domain of *DNMT3A* are associated with an inferior prognosis. *DNMT3A* mutations in GATA3-PTCLs tend to distribute along the N-terminal and PWWP domains. In canine CD4+ PTCL, 36% of *DNM3A* mutations occurred in the N-terminal region, 18% occurred in the DNA methylase domain, and 45% occurred between

domains. In human PTCL-NOS, mutations in *TP53* typically occur in the DNA binding domain, and 44% of canine CD4+ PTCL *TP53* variants were also distributed along this domain.

While these findings indicate some congruity between the mutational profiles of canine $CD4^+$ PTCL and human PTCL-NOS, with similar genes and oncogenic signaling pathways affected, there are limitations to variant calling in RNA-seq data. RNA-seq data is limited to the transcriptome, while DNA sequencing techniques like whole exome sequencing offer broader insight into variations within the larger genome. Additionally, whole exome sequencing allows evaluation of copy number aberrations, which would offer insight into whether amplifications of *STAT3* and *MYC* seen in human GATA3-PTCL are also present in our canine tumors, for example. Another limitation of variant calling both in RNA-seq data and conventional short read whole exome sequencing data is the restriction of these techniques to small variants. For comprehensive deciphering of $CD4^+$ PTCL's mutational profile, long-read whole genome sequencing for large structural variants like chromosomal translocations and large insertions/deletions would also be an essential component.

Conclusions

Our findings indicate that canine $CD4^+$ PTCL is a molecularly comparable disease to human PTCL-NOS with a similar gene expression and mutational profile. Given the increased expression of *GATA3*, evidence of reduced PTEN activity, enrichment for PI3K/AKT/mTOR signaling, *in vitro* dependence on this pathway for tumor cell survival and proliferation, and prevalence of *TP53* mutations, canine $CD4^+$ PTCL appears to most closely resemble the GATA3-PTCL subtype of human PTCL-NOS, which may allow for a more specific application of the proposed canine model for this disease. While we have demonstrated the dependence of canine $CD4^+$ PTCL on PI3K for survival and proliferation, future studies assessing the dependence of these tumors on more downstream effectors within this pathway are warranted to refine the role of PI3K/AKT signaling

pathway in canine CD4$^+$ PTCL. Additionally, variant calling in RNA-seq data has several limitations, and additional whole exome sequencing and long-read whole genome sequencing are important next steps to comprehensively characterize the mutational profile of canine CD4$^+$ PTCL.

REFERENCES

1. Matutes, E. The 2017 WHO update on mature T- and natural killer (NK) cell neoplasms. *Int. J. Lab. Hematol.* 40, 97–103 (2018).

2. Oluwasanjo, A. *et al.* Peripheral T-Cell Lymphoma, not Otherwise Specified (PTCL-NOS). in *T-Cell and NK-Cell Lymphomas: From Biology to Novel Therapies* (eds. Querfeld, C., Zain, J. & Rosen, S. T.) 83–98 (Springer International Publishing, Cham, 2019). doi:10.1007/978-3-319-99716-2_4.

3. Iqbal, J. *et al.* Gene expression signatures delineate biological and prognostic subgroups in peripheral T-cell lymphoma. *Blood* 123, 2915–2923 (2014).

4. Ho, I.-C., Tai, T.-S. & Pai, S.-Y. GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation. *Nat. Rev. Immunol.* 9, 125–135 (2009).

5. Heavican, T. B. *et al.* Genetic drivers of oncogenic pathways in molecular subgroups of peripheral T-cell lymphoma. *Blood* 133, 1664–1676 (2019).

6. Valli, V. E. *et al.* Classification of canine malignant lymphomas according to the World Health Organization criteria. *Vet. Pathol.* 48, 198–211 (2011).

7. Seelig, D. M. *et al.* Canine T-Zone Lymphoma: Unique Immunophenotypic Features, Outcome, and Population Characteristics. *J. Vet. Intern. Med.* 28, 878–886 (2014).

8. Harris, L. J. *et al.* Canine CD4+ T-cell lymphoma identified by flow cytometry exhibits a consistent histomorphology and gene expression profile. *Vet. Comp. Oncol.* 17, 253–264 (2019).

9. Harris, L. J. *et al.* Clinical features of canine nodal T-cell lymphomas classified as CD8+ or CD4−CD8− by flow cytometry. *Vet. Comp. Oncol.* 18, 416–427 (2020).

10. Rebhun, R. B. *et al.* CHOP chemotherapy for the treatment of canine multicentric T-cell lymphoma. *Vet. Comp. Oncol.* 9, 38–44 (2011).

11. Angelo, G., Cronin, K. & Keys, D. Comparison of combination l-asparaginase plus CHOP or modified MOPP treatment protocols in dogs with multi-centric T-cell or hypercalcaemic lymphoma. *J. Small Anim. Pract.* 60, 430–437 (2019).

12. Seelig, D. M., Avery, A. C., Ehrhart, E. J. & Linden, M. A. The Comparative Diagnostic Features of Canine and Human Lymphoma. *Vet. Sci.* 3, (2016).

13. McDonald, J. T. *et al.* Comparative oncology DNA sequencing of canine T cell lymphoma via human hotspot panel. *Oncotarget* 9, 22693–22702 (2018).

14. Elvers, I. *et al.* Exome sequencing of lymphomas from three dog breeds reveals somatic mutation patterns reflecting genetic background. *Genome Res.* 25, 1634–1645 (2015).

15. Cobbold, S. & Metcalfe, S. Monoclonal antibodies that define canine homologues of human CD antigens: Summary of the First International Canine Leukocyte Antigen Workshop (CLAW). *Tissue Antigens* 43, 137–154 (1994).

16. Abrams, V. K. *et al.* A novel monoclonal antibody specific for canine CD25 (P4A10): Selection and evaluation of canine Tregs. *Vet. Immunol. Immunopathol.* 135, 257–265 (2010).

17. Rout, E. D. *et al.* Clinical outcome and prognostic factors in dogs with B-cell chronic lymphocytic leukemia: A retrospective study. *J. Vet. Intern. Med.* 35, 1918–1928 (2021).

18. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890 (2018).

19. Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Res.* 48, D682–D688 (2020).

20. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21 (2013).

21. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930 (2014).

22. Cunningham, F. *et al.* Ensembl 2022. *Nucleic Acids Res.* 50, D988–D995 (2022).

23. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 32, 3047–3048 (2016).

24. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550 (2014).

25. Blighe, K., Rana, S. & Lewis, M. EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling. *R package version 1.14.0* https://github.com/kevinblighe/EnhancedVolcano.

26. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

27. Kassambara, A. *Ggpubr: 'ggplot2' Based Publication Ready Plots*. (2023).

28. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* 16, 284–287 (2012).

29. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 102, 15545–15550 (2005).

30. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740 (2011).

31. Piccaluga, P. P. *et al.* Gene expression analysis of peripheral T cell lymphoma, unspecified, reveals distinct profiles and new potential therapeutic targets. *J. Clin. Invest.* 117, 823–834 (2007).

32. Etebari, M. *et al.* Transcriptional Analysis of Lennert Lymphoma Reveals a Unique Profile and Identifies Novel Therapeutic Targets. *Front. Genet.* 10, (2019).

33. Barrett, T. *et al.* NCBI GEO: archive for functional genomics data sets—update. *Nucleic Acids Res.* 41, D991–D995 (2013).

34. Van der Auwera, G. A. *et al.* From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Protoc. Bioinforma.* 43, 11.10.1-11.10.33 (2013).

35. Krueger, F. *et al.* FelixKrueger/TrimGalore: v0.6.10 - add default decompression path. Zenodo https://doi.org/10.5281/zenodo.7598955 (2023).

36. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv* 1303, (2013).

37. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* 17, 122 (2016).

38. Kandoth, C. *et al.* mskcc/vcf2maf: vcf2maf v1.6.16. Zenodo https://doi.org/10.5281/zenodo.1185418 (2018).

39. Calling variants in RNAseq. *GATK development team* https://github.com/broadgsa/gatk/blob/master/doc_archive/methods/Calling_variants_in_RNAseq.md.

40. Mayakonda, A., Lin, D.-C., Assenov, Y., Plass, C. & Koeffler, H. P. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 28, 1747–1756 (2018).

41. Chakravarty, D. *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO Precis. Oncol.* 1–16 (2017) doi:10.1200/PO.17.00011.

42. Suehnholz, S. P. *et al.* Quantifying the Expanding Landscape of Clinical Actionability for Patients with Cancer. *Cancer Discov.* 14, 49–65 (2024).

43. The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res.* 53, D609–D617 (2025).

44. Avery, P. R. *et al.* Flow Cytometric Characterization and Clinical Outcome of CD4+ T-Cell Lymphoma in Dogs: 67 Cases. *J. Vet. Intern. Med.* 28, 538–546 (2014).

45. Stein, H. *et al.* CD30+ anaplastic large cell lymphoma: a review of its histopathologic, genetic, and clinical features. *Blood* 96, 3681–3695 (2000).

46. Martens, M. *et al.* WikiPathways: connecting communities. *Nucleic Acids Res.* 49, D613–D621 (2021).

47. Majumder, P. K. *et al.* mTOR inhibition reverses Akt-dependent prostate intraepithelial neoplasia through regulation of apoptotic and HIF-1-dependent pathways. *Nat. Med.* 10, 594–601 (2004).

48. Vivanco, I. *et al.* Identification of the JNK Signaling Pathway as a Functional Target of the Tumor Suppressor PTEN. *Cancer Cell* 11, 555–569 (2007).

49. Herek, T. A. *et al.* DNMT3A mutations define a unique biological and prognostic subgroup associated with cytotoxic T cells in PTCL-NOS. *Blood* 140, 1278–1290 (2022).

50. Schatz, J. H. *et al.* Targeted mutational profiling of peripheral T-cell lymphoma not otherwise specified highlights new mechanisms in a heterogeneous pathogenesis. *Leukemia* 29, 237–241 (2015).

51. Huang, Y.-H. *et al.* Genomic and transcriptomic profiling of peripheral T cell lymphoma reveals distinct molecular and microenvironment subtypes. *Cell Rep. Med.* 5, (2024).

52. Das, S., Dixon, J. E. & Cho, W. Membrane-binding and activation mechanism of PTEN. *Proc. Natl. Acad. Sci.* 100, 7491–7496 (2003).

53. Owens, E. *et al.* The gene expression profile and cell of origin of canine peripheral T-cell lymphoma. *BMC Cancer* 24, 18 (2024).

54. Datta, S. R. *et al.* Akt Phosphorylation of BAD Couples Survival Signals to the Cell-Intrinsic Death Machinery. *Cell* 91, 231–241 (1997).

55. Biggs, W. H., Meisenhelder, J., Hunter, T., Cavenee, W. K. & Arden, K. C. Protein kinase B/Akt-mediated phosphorylation promotes nuclear exclusion of the winged helix transcription factor FKHR1. *Proc. Natl. Acad. Sci.* 96, 7421–7426 (1999).

56. Averous, J., Fonseca, B. D. & Proud, C. G. Regulation of cyclin D1 expression by mTORC1 signaling requires eukaryotic initiation factor 4E-binding protein 1. *Oncogene* 27, 1106–1113 (2008).

57. Grewe, M., Gansauge, F., Schmid, R. M., Adler, G. & Seufferlein, T. Regulation of cell growth and cyclin D1 expression by the constitutively active FRAP-p70s6K pathway in human pancreatic cancer cells. *Cancer Res.* 59, 3581–3587 (1999).

58. Gottlob, K. *et al.* Inhibition of early apoptotic events by Akt/PKB is dependent on the first committed step of glycolysis and mitochondrial hexokinase. *Genes Dev.* 15, 1406–1418 (2001).

59. Deprez, J., Vertommen, D., Alessi, D. R., Hue, L. & Rider, M. H. Phosphorylation and activation of heart 6-phosphofructo-2-kinase by protein kinase B and other protein kinases of the insulin signaling cascades. *J. Biol. Chem.* 272, 17269–17275 (1997).

60. Ng, Y., Ramm, G., Lopez, J. A. & James, D. E. Rapid Activation of Akt2 Is Sufficient to Stimulate GLUT4 Translocation in 3T3-L1 Adipocytes. *Cell Metab.* 7, 348–356 (2008).

61. Rathmell, J. C. *et al.* Akt-Directed Glucose Metabolism Can Prevent Bax Conformation Change and Promote Growth Factor-Independent Survival. *Mol. Cell. Biol.* 23, 7315–7328 (2003).

62. Liberti, M. V. & Locasale, J. W. The Warburg Effect: How Does it Benefit Cancer Cells? *Trends Biochem. Sci.* 41, 211–218 (2016).

63. Mulvey, E. & Ruan, J. Biomarker-driven management strategies for peripheral T cell lymphoma. *J. Hematol. Oncol.J Hematol Oncol* 13, 59 (2020).

64. Geng, X. *et al.* GATA-3 is a proto-oncogene in T-cell lymphoproliferative neoplasms. *Blood Cancer J.* 12, 149 (2022).

65. Lemonnier, F. *et al.* Recurrent TET2 mutations in peripheral T-cell lymphomas correlate with TFH-like features and adverse clinical parameters. *Blood* 120, 1466–1469 (2012).

66. Quivoron, C. *et al.* TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. *Cancer Cell* 20, 25–38 (2011).

CHAPTER TWO: The cell of origin of canine CD4$^+$ peripheral T-cell lymphoma[2]

Summary

As established in Chapter One, canine CD4$^+$ PTCL and human PTCL-NOS are molecularly similar diseases and therefore may arise from a similar cell of origin. In humans, the cell of origin of PTCL-NOS remains a subject of ongoing speculation, and investigations into the cell of origin of canine PTCL are limited. In this study, we sought to gain insight into the cell of origin of canine CD4$^+$ PTCL by utilizing gene expression profiling in combination with surface marker expression by flow cytometry. We also endeavored to curate a single-cell transcriptomic atlas of normal canine lymph node and thymus to better characterize the gene expression profiles of normal thymocytes and T cells at various stages of development in the dog and utilize this atlas for the prediction of the cell of origin of canine CD4$^+$ PTCL.

Two bulk RNA-sequencing (RNA-seq) experiments were performed on two cohorts of PTCL samples and sorted CD4$^+$ and CD8$^+$ lymphocytes from healthy control dogs. The first cohort, hereafter referred to as "Cohort 1," consisted of 33 dogs with either CD4$^+$ (26/33), CD8$^+$ (4/33), or CD4-CD8- (3/33) PTCL as diagnosed by flow cytometry and sorted CD4$^+$ and CD8$^+$ lymphocytes from 8 healthy control dogs. The second cohort, hereafter referred to as "Cohort 2," consisted of 96 dogs with CD4$^+$ PTCL diagnosed by flow cytometry, sorted CD4$^+$ nodal lymphocytes from 5 healthy control dogs, and sorted CD4$^+$ thymocytes from 2 healthy control dogs.

Following normalization of bulk RNA-seq data, we performed differential gene expression and unsupervised clustering methods. Gene set enrichment analysis (GSEA) and gene set variant analysis (GSVA) were performed to determine the enrichment of canine CD4$^+$ PTCL for various

[2] Owens E, Harris L, Harris A, Yoshimoto J, Burnett R, Avery A. The gene expression profile and cell of origin of canine peripheral T-cell lymphoma. *BMC Cancer*. 2024 Jan 2;24(1):18. doi: 10.1186/s12885-023-11762-w.

human and murine T-cell and thymocyte gene signatures, as well as gene signatures of normal canine thymocytes and T cells at various stages of development derived from a single-cell transcriptomic atlas of normal canine thymus and lymph node samples.

Single-cell RNA-seq was performed on 3 samples of thymus and 3 samples of lymph node from 3 normal, healthy, <1-year-old dogs for construction of a single-cell transcriptomic atlas of the normal canine T-cell lineage. After normalization of single-cell RNA-seq data, uniform manifold approximation and projection (UMAP) was used to identify clusters based on global gene expression profile. Differential expression analysis between all clusters was then performed to identify features that characterized each cluster, which were then correlated to known biologic cell types. Once cell types were annotated, lymph node and thymus atlases were integrated for trajectory analysis to evaluate how gene expression changes across T-cell development. The gene signatures associated with various stages of thymocyte and T-cell development were then extracted for enrichment analyses using bulk RNA-seq data from Cohort 2.

Canine $CD4^+$ PTCL was enriched for thymic precursor gene signatures, exhibited increased expression of markers of immaturity (*CD34, KIT*, *DNTT,* and *CCR9*), and downregulated genes associated with the T-cell receptor, MHC class II associated genes (*DLA-DQA1, DLA-DRA, HLA-DQB1,* and *HLA-DQB2*), and CD25. Additionally, a subset of canine $CD4^+$ PTCL was enriched for gene signatures associated with human and murine thymocytes and immature T-cell subsets, and for gene signatures associated with immature thymocytes in our canine transcriptomic atlas of normal lymph node and thymus.

In conclusion, a subset of canine $CD4^+$ PTCL may originate from a thymic precursor cell of origin. Given the molecular similarities of canine $CD4^+$ PTCL and human PTCL-NOS, this could challenge the more conventionally posited mature T-helper cell origin of PTCL-NOS.

Introduction

Hematopoietic malignancies like lymphoma and leukemia arise when a normal lymphocyte or lymphoid progenitor cell acquires mutations that initiate malignant transformation, and this originally transformed cell is referred to as the *cell of origin* for these malignancies. Identifying the cell of origin is crucial for understanding the pathogenesis of these cancers, which informs how we can better predict, prevent, and treat them. Identifying the cell of origin of hematopoietic tumors is made challenging by the complex and plastic process of lymphocyte maturation, which spans multiple hematopoietic organs and involves several critical junctures where differentiation into distinct phenotypes is dictated by myriad competing transcription factors and signaling proteins. Additionally, there is increasing evidence that even in their terminally differentiated states, these lymphocyte phenotypes are not fixed. Lineage tracing studies in mice using Cre-Lox recombination systems to track transcription factor expression found that terminally differentiated $CD4^+$ T cells alter their phenotype and cytokine production over the course of their lifespan.[1,2] This plasticity appears to extend to malignantly transformed cells as well, as demonstrated by one study showing that modification of PTCL cell culture conditions altered surface marker and transcription factor expression *in vitro*.[3]

As we demonstrated in Chapter One, canine $CD4^+$ PTCL and human PTCL-NOS are molecularly similar diseases and therefore may arise from a similar cell of origin. While the recently identified TBX21-PTCL and GATA3-PTCL subgroups in human PTCL-NOS invited suppositions that these tumors are derived from their normal Th1 and Th2 immune cell counterparts,[4] the well-established plasticity of T-helper cell differentiation and the resistance of mature T-cells to oncogenic transformation in mice[5] raises an alternative possibility: that these tumors may arise from an immature precursor that is capable of differentiating into a more mature phenotype. Furthermore, *GATA3* has been shown to function as a proto-oncogene and be expressed in many T-

cell tumor types, including precursor neoplasms, independent of the expression of other Th2-specific genes.[6] Investigation into the cell of origin of canine PTCL may help offer additional insight into the cell of origin in human PTCL-NOS.

In one study of 6 dogs with $CD4^+$ PTCL, Harris et al.[7] speculated that the cell of origin could be a thymic precursor cell rather than a mature, activated T cell due to the frequent presentation of these tumors with a mediastinal mass and the low surface expression of CD25 and MHC class II by $CD4^+$ PTCL cells. However, investigations into the cell of origin of canine $CD4^+$ PTCL using RNA-sequencing to compare its gene expression profile to that of normal lymphocytes and thymocytes at various stages of development are lacking. Recently, an RNA-sequencing study comparing gene expression patterns of canine T-zone lymphoma cells to gene signatures of human and murine naïve and activated T-cell subsets was successful in aiding the identification of a mature, previously activated αβ T cell as the cell of origin for canine T-zone lymphoma.[8] A similar approach may therefore have utility in identifying the cell of origin of canine $CD4^+$ PTCL.

A major limitation to utilizing gene expression profiling to probe the cell of origin of $CD4^+$ PTCL is the lack of reference gene signatures for normal canine thymocytes and lymphocytes. While we can still infer a great deal from the more widely available human and murine data, we know that there are species-specific differences in immunology that complicate broad generalizations about the molecular features of immune cells between species. For example, in mice, expression of MHC class II molecules is restricted to antigen presenting cells and is not expressed by T cells, but in dogs, MHC class II expression is a feature of all peripheral T cells, although its expression in the neonatal thymus and on resting T cells is weaker than on activated T cells.[9,10] Given that the canine immune system has not been studied as exhaustively as the human and murine immune systems, there are likely many other deviations that complicate inference of cell of origin from human and murine immune cell counterparts.

One method that could overcome the limited availability of gene sets describing normal canine hematopoietic cells is single-cell RNA-seq. In bulk RNA-seq, RNA is extracted from the collective cells that comprise the sample, and average gene expression across all cells is measured. This method is most informative for relatively pure samples or homogeneous tumors, where we can assume gene expression patterns will be similar across all cells in the sample. In contrast, during single-cell RNA-seq on a 10X Genomics platform, individual cells are suspended in gel beads coated with oligonucleotides containing a 16-nucleotide barcode that tracks which cell each RNA molecule originates from, and a unique molecular identifier (UMI) that acts as a fingerprint for each transcript to enable elimination of any downstream PCR duplicates. Thus, every RNA molecule that originates from the same cell will have the same barcode, but its own UMI. Cells can then be clustered according to similarities in their global gene expression profile, and the biologic cell type corresponding to these clusters predicted based on their differential expression of canonical markers. In summary, single-cell RNA-seq offers information about gene expression patterns across individual cell types in a heterogeneous sample, which would allow us to characterize the gene expression profiles of multiple populations of thymocytes, naïve, and mature T-cell subsets from samples of thymus and lymph node. These gene expression profiles could then be compared to $CD4^+$ PTCL cells to aid in the identification of the likely cell of origin of these tumors.

Additionally, with single-cell RNA-seq data, we can perform pseudotemporal ordering to track how thymocytes and T cells are changing as they progress through different stages of maturation and differentiation. In this analysis, the gene expression profile of each cell gets represented as a point in a high-dimensional Euclidean space where all genes have their own dimension. For ease of visualization and interpretation, this high-dimensional space is reduced to a low-dimensional space that still preserves the relationships between cell populations using independent component analysis. Then, a minimum spanning tree is overlaid, and the longest path

of transcriptionally similar cells is identified and used to produce a “trajectory” of an individual cell’s progress through differentiation. After the longest sequence of similar cells is identified, cells not on this path are examined for construction of alternative trajectories that represent branches of cell fate decisions. These alternative trajectories are then also ordered and connected to the main trajectory. The end result is the ordering of single-cell gene expression profiles in *pseudotime*, or a quantitiative emulation of their progression through a biologic state like differentiation. Once cells are ordered in pseudotime, we can find genes that are dynamically expressed as cells progress from one stage of differentiation to the other. If we then recognize expression of these key regulators in $CD4^+$ PTCL, it may offer insight into the developmental stage of the corresponding cell of origin of these tumors.

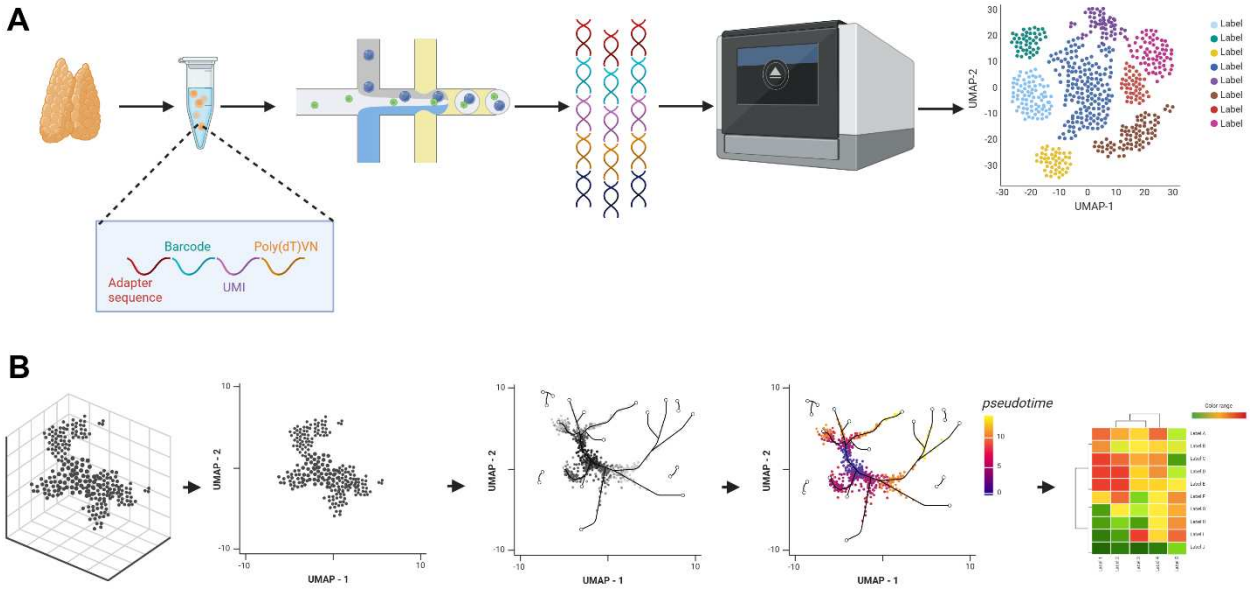


Figure 17: Overview of single-cell RNA-seq and pseudotemporal ordering. (**A**) Individual cells are suspended in gel beads coated with oligonucleotides containing a barcode that tracks which cell each RNA molecule originates from, and a unique molecular identifier (UMI) that acts as a fingerprint for each transcript. Cells then move through a microfluidic system to generate gel beads-in-emulsions (GEMs), each containing a single cell. Once cells are lysed and the gel bead dissolved, the 3’ end of RNA binds poly-dT sequences that coated the gel bead. RNA is then reverse transcribed to form cDNA libraries for next-generation sequencing. Barcoding transcripts with the cell they originated from allows us to measure gene expression for each individual cell in a sample. Cells are then clustered based on similarities in gene expression profiles and annotated with their corresponding biologic cell type by evaluating differential expression of key marker genes. (**B**) Cells are represented as points in a high-dimensional expression space where every gene is represented by a dimension. This dimensionality is then reduced via independent component analysis, and a minimum spanning tree overlaid. The longest path through the minimum spanning tree represents the longest sequence of transcriptionally similar cells. Cells that are separate from this longest sequence are ordered as alternative trajectories and represent branches of cell fate decisions. Once single-cell gene expression profiles are ordered in pseudotime, differential gene expression analysis can identify genes that change as cells move between stages of differentiation. Created in BioRender. Owens, E. (2025) https://BioRender.com/u85m736.

Therefore, the goals of this study were to 1) perform flow cytometry to evaluate differences in expression of surface markers associated with T-cell activation and maturity between canine $CD4^+$ PTCL, control canine lymph nodes, and control canine thymuses; 2) use bulk RNA-sequencing data to evaluate the expression of genes associated with T-cell immaturity and maturity and compare the gene expression profile of $CD4^+$ PTCL to gene signatures across human and murine T-cell development; 3) develop a single-cell transcriptomic atlas of normal canine lymph node and thymus and perform pseudotemporal trajectory analysis to identify key genes that change as cells move through differentiation states; and 4) utilize this atlas to compare the gene expression profile of canine $CD4^+$ PTCL to various stages of canine thymocyte and T-cell development for cell of origin inference.

Material and Methods

Sample collection

The PTCL, sorted $CD4^+$ and $CD8^+$ lymphocyte, and sorted $CD4^+$ thymocyte samples utilized for bulk RNA-seq in this study were the same as those previously described for Cohort 1 and Cohort 2 in the Materials & Methods section of Chapter One. Sample collection and library preparation for single-cell RNA-seq was performed by another member of our lab, Dr. Adam Harris. For single-cell RNA-seq, single cell suspensions were prepared from fresh thymus and lymph node of three reportedly healthy 7- to 9-month-old mixed breed dogs designated for IACUC-approved medical device research. Red blood cells were lysed with ammonium-chloride-potassium lysis buffer, and composition of single-cell suspensions was verified by cytology and immunophenotyping by flow cytometry using a Beckman Coulter three laser Gallios flow cytometer or a Sony ID7000 spectral flow cytometer. Propridium iodide was used to perform viability checks. Neutrophil proportions were determined by enumerating cells with a $CD4^+CD3^-CD5^-CD18^+$ immunophenotype by flow

cytometry. All whole thymus and lymph node samples composed of >70% live cells and <25% mature neutrophils were used for single-cell RNA-seq.

RNA-sequencing

For bulk RNA-sequencing of Cohort 1 and Cohort 2, RNA was extracted, purified, and sequenced as previously described in the Materials & Methods section of Chapter One. For single-cell RNA-sequencing of normal canine thymus and lymph node, cells were resuspended in phosphate-buffered saline and 0.04% molecular grade bovine serum albumin. 5,000 cells were targeted for capture using a Chromium iX instrument (10X Genomics). Standard Illumina libraries were generated following manufacturer-recommended protocols with a Chromium Next GEM Single Cell 3' Kit v3.1 dual index (10X Genomics) following generation of gel bead-in emulsions (GEMs). Libraries were sequenced by Novogene Corporation Inc (Sacramento, CA) on an Illumina NovaSeq 6000 sequencer targeting 100,000 reads/cell with 150 bp paired-end sequencing.

Data analysis

Bulk RNA-seq reads for Cohort 1 and 2 were preprocessed, aligned to the CanFam3.1 reference genome, tabulated, normalized, and analyzed for differential gene expression as previously described in the Materials & Methods section of Chapter One. Box plots of DESeq2 normalized count data were generated using ggplot2[11] (version 3.5.1), and means of normalized counts were compared by Wilcoxon signed-rank test using the ggpubr[12] (version 0.6.0) package. GSEA was performed on bulk RNA-seq data with clusterProfiler[13] (version 4.12.6) using the collective gene sets available through the Molecular Signatures Database (MSigDB)[14] and curated gene sets containing the top 250 upregulated and downregulated genes (ranked by log2 fold change) in human CD4 SP infant thymocytes compared to $CD4^+$ T-cells in adult and infant blood obtained from supplementary materials in Helgeland et al.[15] The Wald test statistic was selected as

the ranking metric for the list of differentially expressed genes in canine CD4$^+$ PTCL (as obtained by DESeq2) for its consideration of p-value in conjunction with log2 fold change. clusterProfiler GSEA parameters were set to perform 10,000 permutations and enforce a p-value cutoff of 0.05 following correction using the Benjamini-Hochberg procedure. Gene set variation analysis (GSVA) was conducted using GSVA[16] (version 1.52.3) to calculate the enrichment of individual CD4$^+$ PTCL and control CD4$^+$ lymphocyte samples for the expression profiles of various stages of thymocyte development (GSE1460), resting and activated human immune cells (GSE22886), and murine T-helper cell subsets (GSE14308) available from MSigDB, as well as curated gene sets containing the top 250 upregulated and downregulated genes (ranked by log2 fold change) in human CD4 SP infant thymocytes compared to CD4$^+$ T-cells in adult and infant blood obtained from supplementary materials in Helgeland et al.[15] Variance stabilized transformed read count data from canine CD4$^+$ PTCL samples and control CD4$^+$ lymphocytes (as performed by DESeq2) was used for this analysis. Differential expression analysis to determine significantly enriched gene sets was subsequently performed on GSVA enrichment scores with limma[17] (version 3.60.4) using default parameters. GSVA results were filtered to include only those gene sets found to be significantly differentially expressed (padj <0.05), and hierarchical Ward-linkage clustering of GSVA enrichment data based on Spearman's correlation coefficient for samples and the Pearson correlation coefficient for gene sets was performed as recommended in the GSVA vignette. Heatmaps of hierarchical clustering of GSVA enrichment data were generated with the heatmap.2 function in gplots (version 3.2.0).[18] The full bioinformatics pipeline and software details for the bulk RNA-seq analysis have been made available in a public Github repository: https://github.com/pirateyoho/CaninePTCLRNASeq.

A Tukey's multiple comparisons test was performed using GraphPad Prism for Windows, GraphPad Software, San Diego, California USA, www.graphpad.com (version 10.0.2) to evaluate

differences in surface marker expression measured by flow cytometry between canine thymocytes, canine PTCL, and normal canine nodal $CD4^+$ T-cells.

Raw single-cell RNA-seq reads in FASTQ file format were made available to us by Dr. Adam Harris. Raw reads were aligned to the CanFam3.1 reference genome (accessed through Ensembl release 104[19]) and tabulated according to a standard 10X Genomics Cell Ranger (version 7.1.0) pipeline. Count matrices were then imported into R[20] (versions 4.4.0 and 4.4.1) for downstream analysis. Initial quality control and filtering was performed using Seurat (versions 5.1.0 and 5.2.0). Low quality cells with <500 UMIs per cell, <300 or >6000 genes detected per cell, and >20% mitochondrial reads were filtered before downstream analysis. Genes with zero counts in all cells or genes expressed in less than 10 cells were also removed from the data prior to downstream analysis to avoid dramatic reductions in the average expression for a cell. After applying these initial filters, DoubletFinder (version 2.0.4) was used to predict doublets, where multiple cells are mistakenly counted as one cell. The vast majority of predicted doublets overlapped with clusters of cells expressing cell cycle associated genes (Supplemental Figure 5), and we therefore suspected the high amount of transcript in these cells was biologically relevant rather than artifact and opted not to filter these doublet calls from our data.

Data was then processed through the standard preprocessing workflow in Seurat, which includes normalization of the data with the Seurat::NormalizeData() function, identification of the features exhibiting the highest cell-to-cell variation in the dataset using the Seurat::FindVariableFeatures() function, scaling of the data via the Seurat::ScaleData() function, and linear dimensional reduction of the data with a principal component analysis (PCA) using the Seurat::RunPCA() function. Principal components were then ranked based on the percentage of variance explained by each (Supplemental Figure 6), and an elbow plot of these ranked principal components was used to observe how many of the top principal components captured the majority

of the true signal. This number of principal components (10 principal components for lymph node and 10 principal components for thymus) was provided as input for nearest-neighbor graph construction using the Seurat::FindNeighbors() function (Supplemental Figure 6). Cells were then clustered at resolutions 0.1-1.1 using the Seurat::FindClusters() function. To determine the optimal number of clusters for the data, cluster stability at different resolutions was evaluated with clustree (version 0.5.1) (Supplemental Figure 7) and visual inspection of over- or under-clustering following non-linear dimensional reduction with the uniform manifold approximation and projection (UMAP). A clustering resolution of 0.2 and 0.1 were ultimately selected for the thymus and lymph node, respectively, based on these evaluations, and the Seurat::FindClusters() function was re-run at these resolutions. Visual inspection of a range of values for the *n.neighbors* parameter (the number of neighboring points used in local approximations of manifold structure) and *min.dist* parameter (how tightly points are allowed to be compressed together) of the Seurat::RunUMAP() function was then performed, and ultimately n.neighbors = 50 and min.dist = 0.5 was selected for the thymus data, and n.neighbors = 30 and min.dist = 0.3 for the lymph node data.

Clusters were then annotated using a combination of manual evaluation of key marker expression using the Seurat::FindAllMarkers() function, gene set enrichment analysis using clusterProfiler (version 4.14.4), cell cycle scoring with Seurat::CellCycleScoring(), and module scoring using the Seurat::AddModuleScore() function. The gene sets used for these analyses included the top 51 enriched genes for the “DN”, “ETP”, “DP”, “CD4+T”, “CD8+T”, “T(agonist)”, “ILC3”, “CD4+Tmem”, “Treg”, “NK”, “CD8+Tmem”, “NKT”, “Mgk”, “Ery”, “T_DN”, “Mast”, “MEMP”, “HSC”, “NMP”, “B_pro/pre”, “pDC”, “DN(early)”, “DN(P)”, “DN(Q)”, “DP(P)”, “DP(Q)”, “Treg(diff)”, and “Th17”, clusters of the Park et al[21] human fetal thymus atlas accessed through the University of California, Santa Cruz (UCSC) Cell Browser (https://cells.ucsc.edu/?ds=fetal-thymus), the top 51 enriched genes for the “Trm_gut_CD8”, “Tnaive/CM_CD4”, “Tregs”, “Pro-B”, “Naïve B cells”, “Pre-

B", "Plasma cells", "Plasmablasts", "Memory B cells", "Erythroid", "pDC", "Megakaryocytes", "Mast cells", "ILC3", "Cycling T&NK", "DC1", "DC2", "Classical monocytes", "Cycling", "Progenitor", "Tem/emra_CD8", "migDC", "Trm_Tgd", "NK_CD56bright_CD16-", "NK_CD16+", "Tgd_CRTAM+", "Trm/em_CD8", "MAIT", "Tnaive/CM_CD4_activated", "Trm_Th1/Th17", "T_CD4/CD8", "Tnaive/CM_CD8", "Teffector/EM_CD4", and "Tfh" clusters of the Conde et al.[22] human single-cell immune cell atlas accessed through the UCSC Cell Browser (https://cells.ucsc.edu/?ds=pan-immune+global), and the top 51 enriched genes for the "Immature B cell", "Naïve B cell", "CD8 gd T cell", "Cycling T cell", "CD34+ Unclassified", "Unclassified DC", "Pre-DC", "PMN-MDSC", "Plasmacytoid DC", "Plasma cell", "NK T cell", "DN T cell", "CD8+ Naïve", "CD4+ Naïve", "CD4+ TCM", "CD4+ T reg", "CD4+, IFN signature", "CD4+ TEM", "CD4+ TEM, Th1 like", "CD4+ TEM, Th17 like", "CD4+ TEM, Th2 like", "CD8+ Memory", "CD8+ Effector", "NK cell", "gd T cell", "Myeloid cDC2", "Monocyte, CD4-", "Monocyte, IFN signature", "Monocyte, CD4+", "M-MDSC", "Basophil", "Neutrophil", "Eosinophil", "Class switched B cell", and "Activated B cell" clusters of the Ammons et al.[23] canine circulating leukocyte atlas accessed through the UCSC Cell Browser (https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy). To improve clarity when referencing these gene sets in figure legends, the original cluster names for these reference atlases were replaced with unique gene set names for our analyses (Supplemental Table 11). In addition to these gene sets, a curated list of gene sets containing genes upregulated by various human and murine T-cell progenitor and thymocyte subsets from the Broad Institute Molecular Signatures Database[14] were also used for cluster annotation by gene set enrichment analysis (Supplemental Table 12). The full list of markers evaluated for cluster annotation and results of these analyses are available on RPubs (Supplemental Table 13).

Thymus and lymph node were initially analyzed and annotated separately, then T-cell populations were subset using Seurat::subset() to exclude any non-T-cell-lineage-associated clusters.

Thymocyte and T-cell populations from both data sets were then merged with Seurat::merge() and integrated with Seurat::IntegrateLayers() and Seurat::JoinLayers() using canonical correlation analysis (CCA) for batch effect correction. After merging and integration of thymus and lymph node data, single-cell RNA-seq counts were re-processed via the standard preprocessing workflow in Seurat described above. 10 principal components were used for nearest-neighbor graph construction, and a clustering resolution of 0.2 was chosen based on evaluation of cluster stability with clustree in combination with subjective evaluation of over- or under-clustering on UMAP plots.

Pseudotemporal trajectory analysis was performed with Monocle 3 (version 1.3.7). First, the Seurat object of merged thymus and lymph node was converted to a celldata set object for Monocle 3 using the as.cell_data_set() function of SeuratWrappers (version 0.4.0). Size factors were then calculated using monocle3::estimate_size_factors() and cells clustered with monocle3::cluster_cells(). Prior to trajectory analysis, dimensionality of the celldata set was reduced with monocle3::reduce_dimension(). Trajectory analysis was then performed with monocle3::learn_graph() and cells ordered according to pseudotime using monocle3::order_cells(). The cluster of early thymic progenitor cells and doublet negative thymocytes was provided to order_cells() as the starting cells. Differential expression analysis to identify the top markers defining each cluster in the celldata set was performed with monocle3::top_markers(). Variable genes were identified using Morans *I* spatial autocorrelation with monocle3::graph_test(), and genes were assigned to modules of similar expression patterns based on Morans *I* values using Louvain community analysis with monocle3::find_gene_modules() and monocle3::aggregate_gene_expression(). Dynamics of gene expression across pseudotime were visualized with monocle3::plot_genes_in_pseudotime(). The full bioinformatics pipeline and software details for the single-cell RNA-seq and pseudotemporal trajectory analysis have been made available in a public Github repository: https://github.com/pirateyoho/CanineSingleCellRNASeq.

Results

Canine $CD4^+$ PTCL downregulates markers of T-cell activation and T-cell receptor signaling

By flow cytometry, canine $CD4^+$ PTCL samples were noted to have significantly lower CD25 ($p < 0.0001$) (Figure 18A) and MHC Class II expression ($p < 0.0001$) (Figure 18B) than $CD4^+$ T-cells from control lymph nodes. Similarly low surface expression of CD25 and MHC Class II was appreciated in normal canine thymocytes by flow cytometry (Figure 18A-B), suggesting that low expression of these surface antigens may be reflective of earlier stages of T-cell maturation. Specifically, concurrent high surface CD3 expression with low surface MHC class II expression, like what was observed in canine $CD4^+$ PTCL cells, was most characteristic of DP and CD4 SP thymocytes in the canine thymus (Figure 18C).

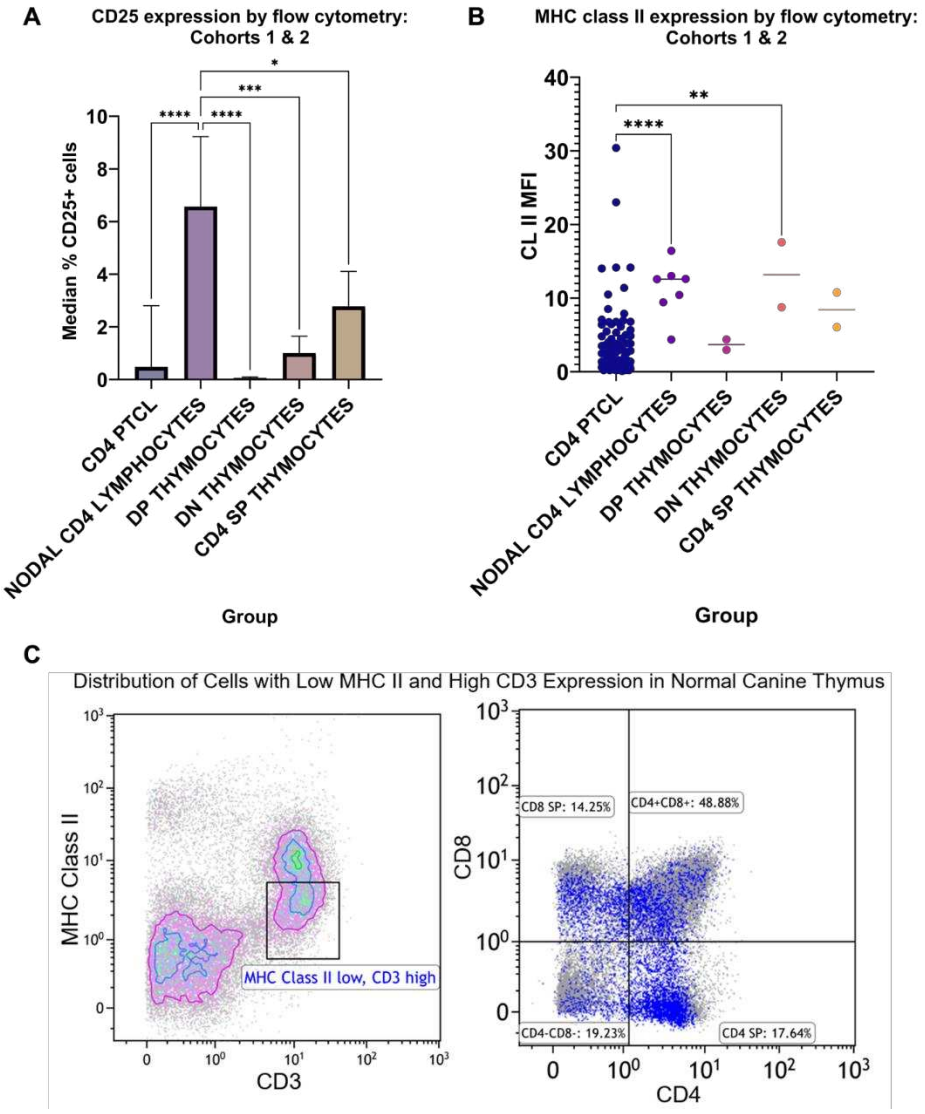


Figure 18: CD25 and MHC class II expression in CD4⁺ PTCL and normal canine thymocytes and CD4⁺ lymphocytes. CD4⁺ PTCL samples exhibit significantly lower CD25 ($p < 0.0001$) (**A**) and MHC Class II expression ($p < 0.0001$) (**B**) compared to CD4⁺ T-cells from control lymph nodes. Normal canine thymocytes demonstrate similarly low expression of these antigens by flow cytometry. (**C**) DP and CD4⁺ SP thymocytes predominantly composed the CD3-high, MHC class II-low population in the thymus that most closely resembles the flow cytometric features of CD4⁺ PTCL.

The low CD25 and MHC class II surface expression in canine CD4⁺ PTCL observed by flow cytometry was corroborated by our gene expression data, which found that canine CD4⁺ PTCL had significantly decreased expression of *IL2RA* and MHC class II gene *HLA-DQB1* in both cohorts compared to control canine CD4⁺ lymphocytes (Figure 19A-B, Table 4). CD4⁺ PTCLs in Cohort 1 additionally demonstrated significantly decreased expression of MHC class II-associated genes *DLA-DQA1, HLA-DQB2,* and *DLA-DRA* (Figure 19A).

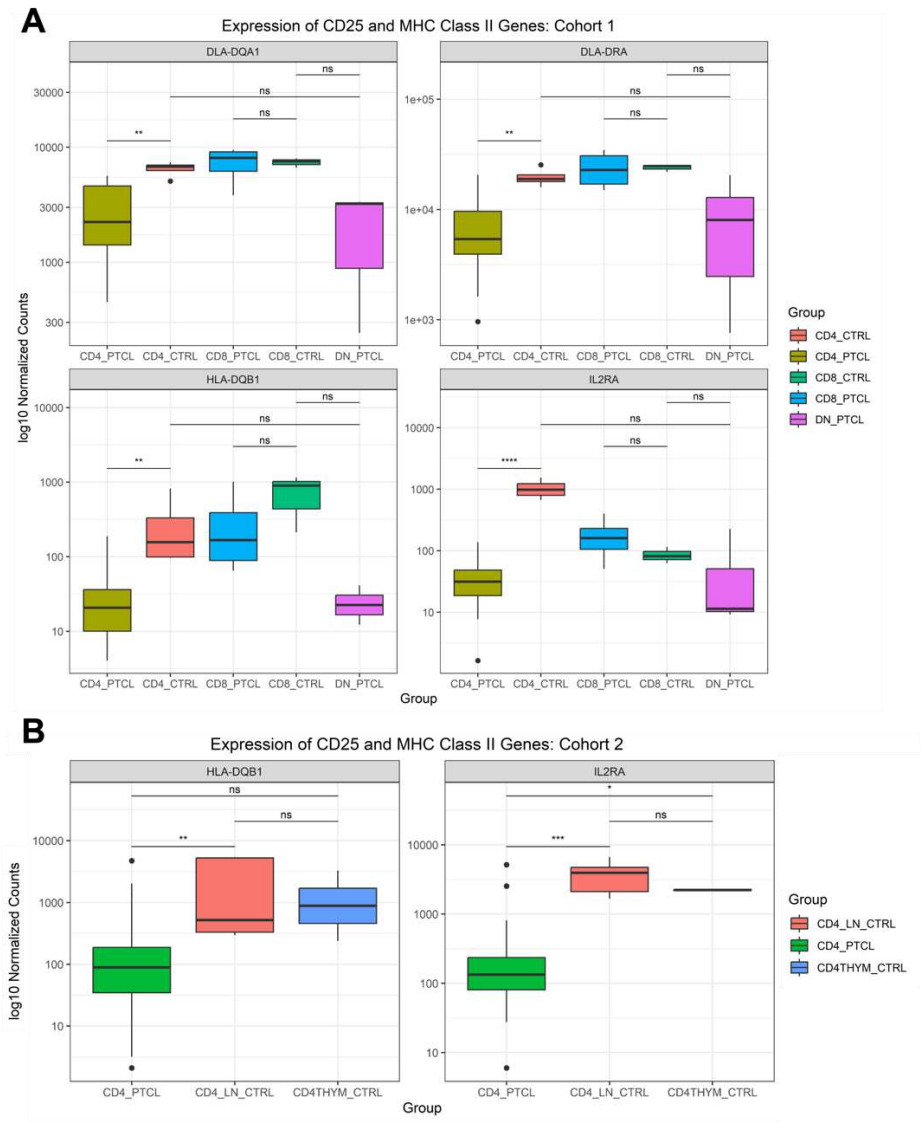
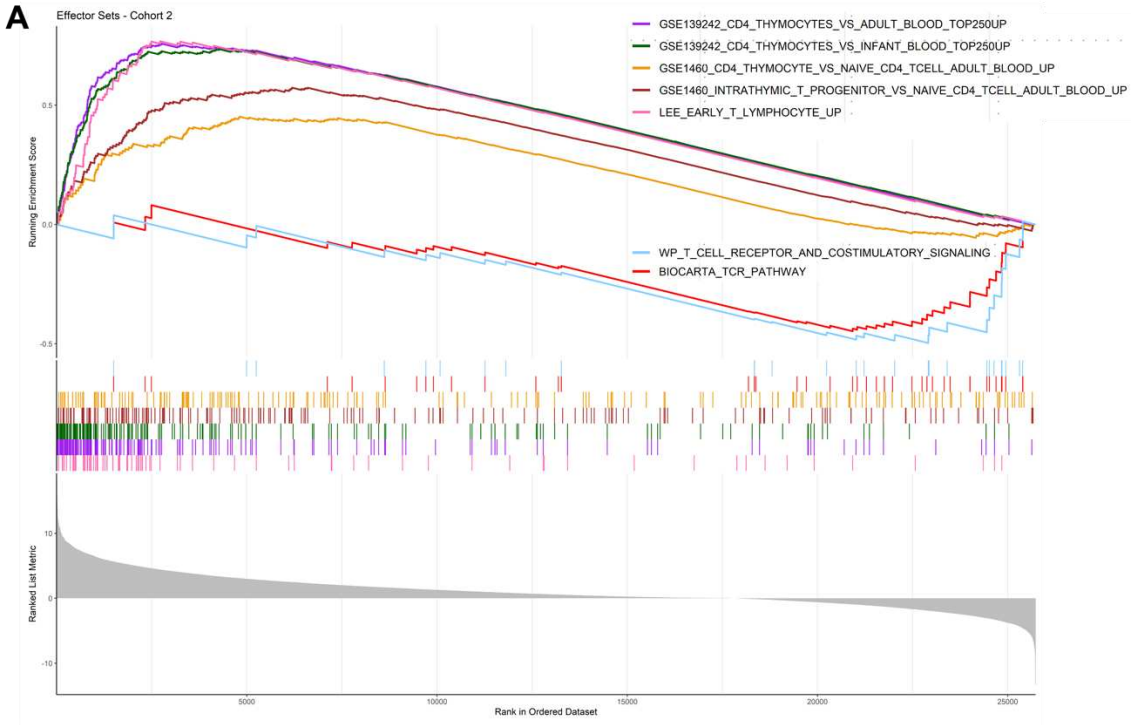


Figure 19: CD25 and MHC class II associated gene expression. A Wilcoxon signed-rank test comparing the means of RNA-seq counts normalized by median of ratios identified significant downregulation of CD25-encoding *IL2RA* and MHC class II-associated genes *DLA-DQA1, DLA-DRA*, and *HLA-DQB1* in Cohort 1 (**A**), and *IL2RA* and *HLA-DQB1* in Cohort 2 (**B**).

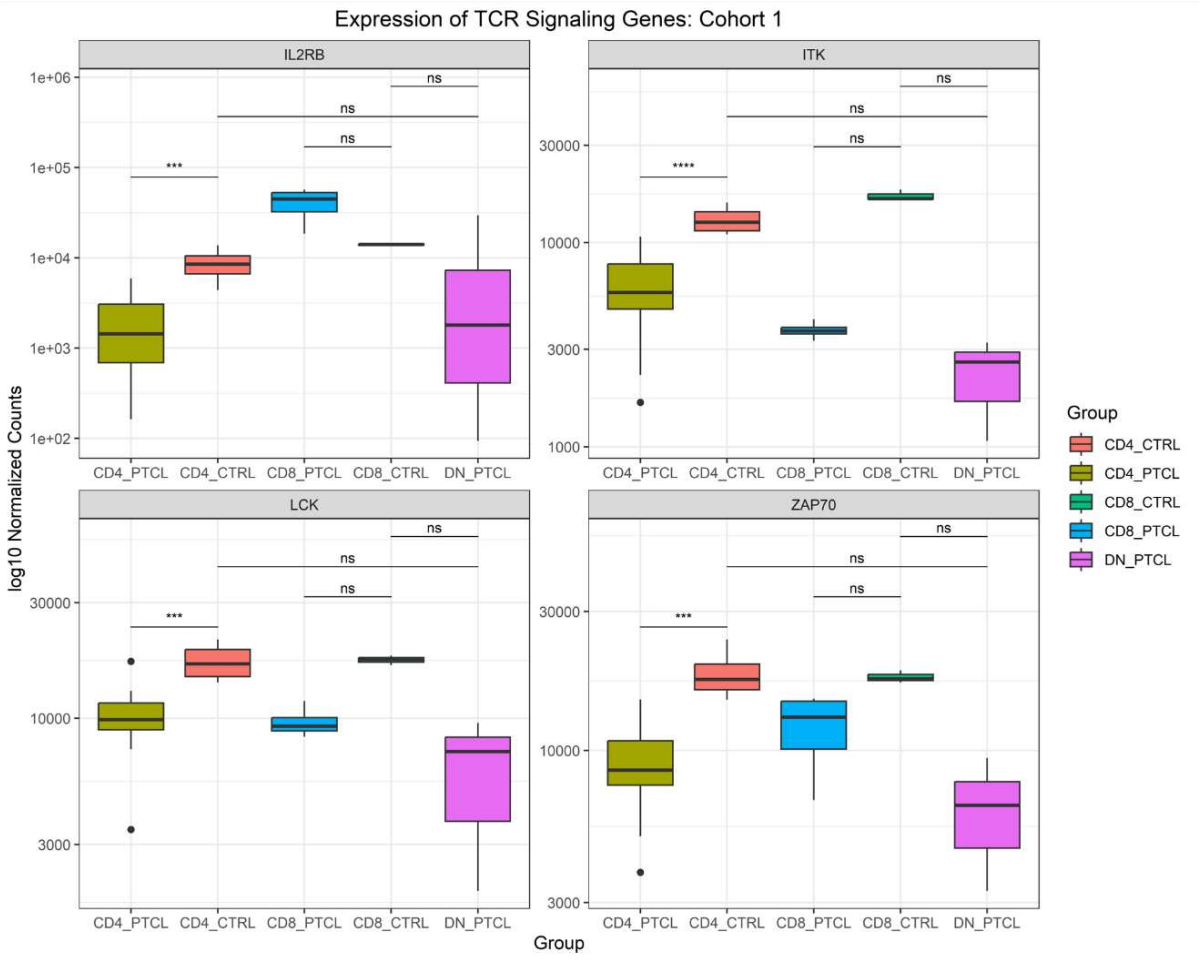
GSEA comparing the gene expression profile of canine CD4$^+$ PTCL to the Biocarta[24] TCR signaling pathway and Wikipathways[25] TCR and co-stimulatory signaling pathway (WP2583) found that canine CD4$^+$ PTCL was negatively enriched for these gene signatures (NES = -1.78, FDR q-value = 0.01; and NES = -1.77, FDR q-value = 0.03, respectively, for Cohort 2) (Figure 20A), indicating that TCR signaling may be downregulated in canine CD4$^+$ PTCL. This was supported by

evidence of significant downregulation of genes associated with TCR signaling, including *LCK, ITK, ZAP70*, and *IL2RB* (Table 4, Figure 20B-C).

Taken together, these findings suggest that canine $CD4^+$ PTCL does not originate from an activated $CD4^+$ T-cell phenotype.



B


Expression of TCR Signaling Genes: Cohort 1
IL2RB
ITK
LCK
ZAP70
log10 Normalized Counts
Group
CD4_PTCL
CD4_CTRL
CD8_PTCL
CD8_CTRL
DN_PTCL
ns
***
****

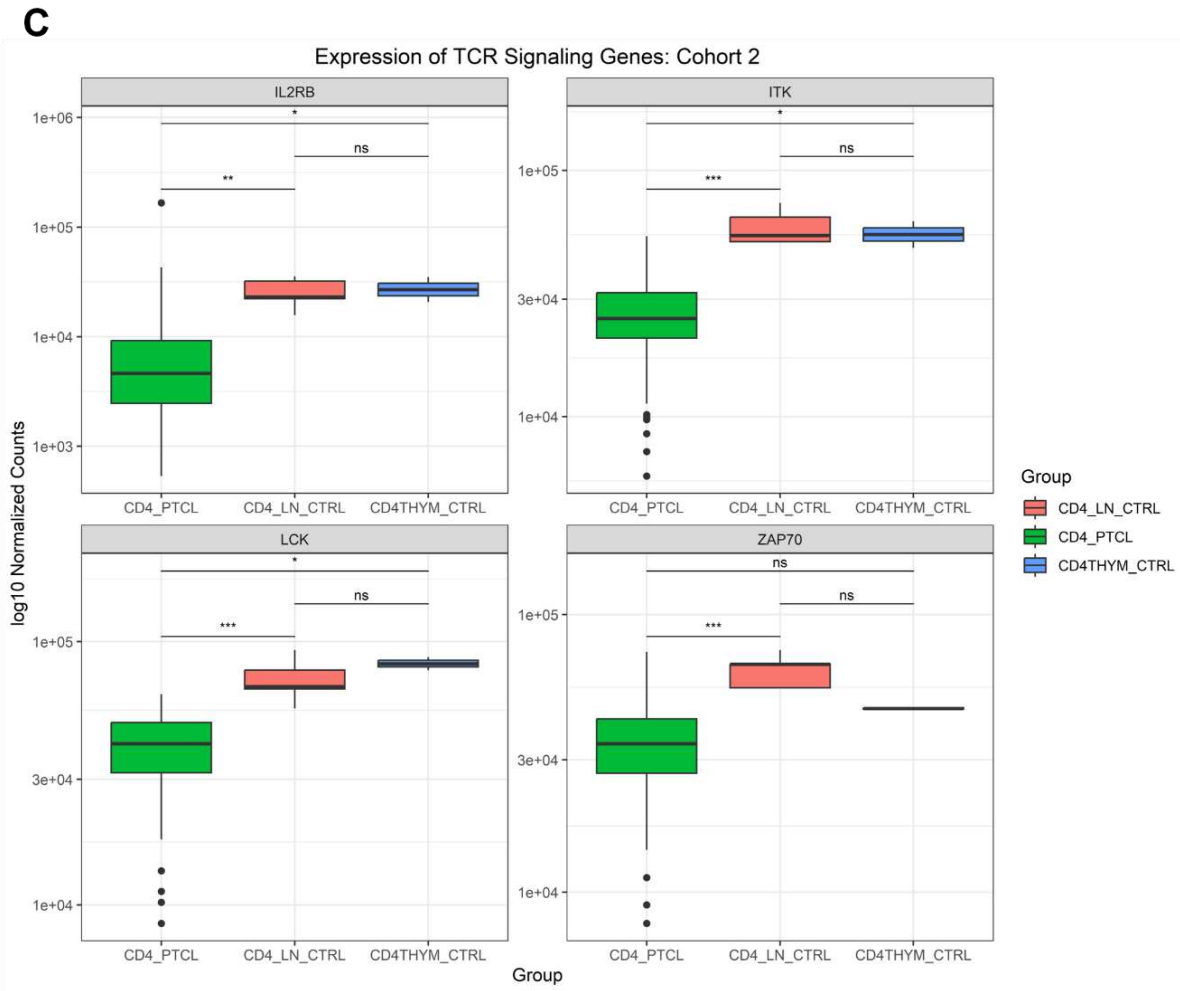


Figure 20: TCR signaling expression and enrichment in CD4⁺ PTCL. (**A**) Purple = enrichment of canine CD4⁺ PTCL for genes upregulated in CD4⁺ SP thymocytes compared to circulating adult CD4⁺ T cells (GSE139242); green = enrichment of canine CD4⁺ PTCL for genes upregulated in CD4 SP thymocytes compared to circulating infant CD4⁺ T-cells (GSE139242); orange = enrichment of canine CD4⁺ PTCL for genes upregulated in CD4 SP thymocytes compared to naïve CD4 T-cells (GSE1460); brown = enrichment of canine CD4⁺ PTCL for genes upregulated in intrathymic T progenitors compared to naïve CD4 T-cells (GSE1460); pink = enrichment of canine CD4⁺ PTCL for genes upregulated in early T-lymphocytes (GSE1460); blue = negative enrichment of canine CD4⁺ PTCL for TCR and costimulatory signaling (Wikipathways WP2583); red = negative enrichment of canine CD4⁺ PTCL for TCR signaling (Biocarta TCR signaling pathway). Results had a Benjamini–Hochberg p-value ≤ 0.05. (**B-C**). A Wilcoxon signed-rank test comparing the means of normalized count data between canine PTCL subtypes and control nodal CD4⁺ and CD8⁺ lymphocytes in Cohort 1 (**B**) and between canine CD4⁺ PTCL and control nodal CD4⁺ lymphocytes and CD4⁺ thymocytes in Cohort 2 (**C**) identified significantly decreased expression of TCR signaling associated genes *IL2RB, ITK, LCK,* and *ZAP70* in CD4⁺ PTCLs.

Table 4: Results of DESeq2 differential expression analysis of *IL2RA*, MHC class II genes, and TCR signaling genes between canine CD4⁺ PTCL and control nodal CD4⁺ T cells.

Cohort 1		
Gene	**Log2 fold change**	**padj**
HLA-DQB1	-3.23	0.0007
DLA-DQA1	-1.22	0.04
DLA-DRA	-1.37	0.03

IL2RA	-4.37	2.01E-11
IL2RB	-2.18	0.006
ITK	-1.08	0.003
LCK	-0.73	0.008
ZAP70	-1.02	0.0001
Cohort 2		
Gene	**Log2 fold change**	**Padj**
HLA-DQB1	-3.13	0.0008
DLA-DQA1	-0.35	0.60
DLA-DRA	-0.97	0.11
IL2RA	-3.78	4.56E-8
IL2RB	-1.44	0.066
ITK	-1.13	9.97E-05
LCK	-0.84	0.001
ZAP70	-0.84	0.002

Canine CD4$^+$ PTCL upregulates genes associated with T-cell immaturity and is enriched for gene signatures associated with earlier stages of human and murine T-cell development

GSEA comparing the gene expression profile of canine CD4$^+$ PTCL to gene sets associated with subpopulations of human thymocytes and circulating naïve CD4$^+$ T cells (GSE1460[26] and GSE139242[15]) revealed significant enrichment of canine CD4$^+$ PTCL for gene signatures associated with early T lymphocytes, intrathymic T progenitor cells, CD4 SP thymocytes, and CD4/CD8 double-positive (DP) thymocytes compared to naïve circulating CD4$^+$ T-cells (Table 5, Figure 20A).

Table 5: Enrichment scores for human thymocyte and circulating naïve CD4$^+$ T cell gene sets. Enrichment analysis was performed with clusterProfiler using the ranked list of differentially expressed genes identified between canine CD4$^+$ PTCL cells and control canine nodal CD4$^+$ lymphocytes.

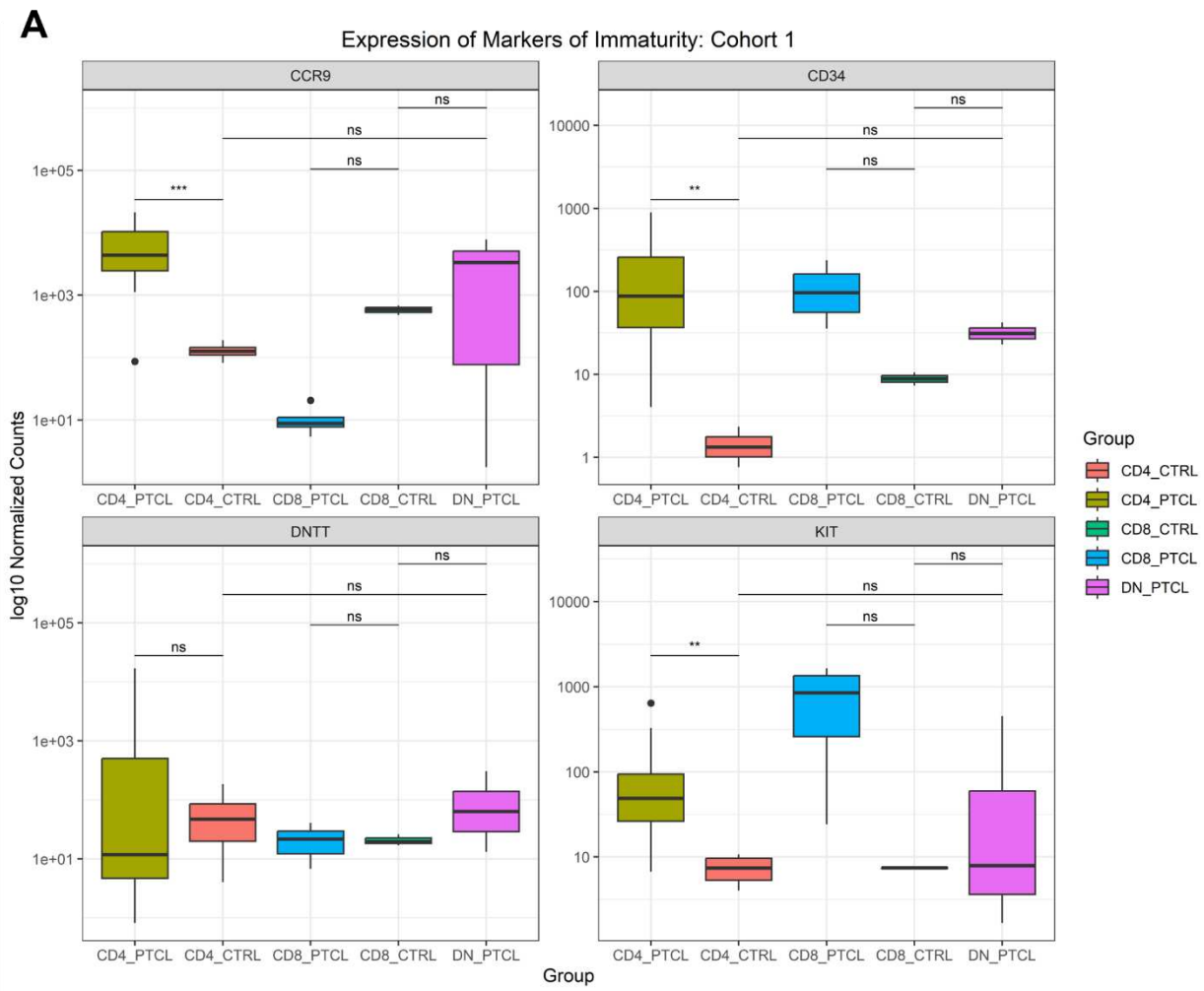
Cohort 1			
Gene Set	**Normalized enrichment score**	**padj**	**FDR**
GSE1460: Early T lymphocyte gene signature	2.66	0.003	0.002
GSE1460: Genes upregulated in intrathymic T-progenitor cells compared to naïve CD4$^+$ T-cells	2.08	0.003	0.002
GSE1460: Genes upregulated in CD4 SP thymocytes compared to naïve CD4$^+$ T-cells	1.58	0.006	0.005
GSE1460: Genes upregulated in CD4/CD8 DP thymocytes compared to naïve CD4$^+$ T-cells	1.51	0.01	0.01
GSE139242: Genes upregulated in infant CD4$^+$ thymocytes compared to CD4$^+$ T-cells in infant blood	2.78	4.25e-10	5.26e-11

GSE139242: Genes upregulated in infant CD4$^+$ thymocytes compared to CD4$^+$ T-cells in adult blood	2.50	4.25e-10	5.26e-11
Cohort 2			
Gene Set	**Normalized enrichment score**	**padj**	**FDR**
GSE1460: Early T lymphocyte gene signature	2.25	1.10E-09	4.91E-10
GSE1460: Genes upregulated in intrathymic T-progenitor cells compared to naïve CD4$^+$ T-cells	1.75	1.38E-09	6.18E-10
GSE1460: Genes upregulated in CD4 SP thymocytes compared to naïve CD4$^+$ T-cells	1.38	0.004	0.002
GSE1460: Genes upregulated in CD4/CD8 DP thymocytes compared to naïve CD4$^+$ T-cells	1.49	0.0001	5.16E-05
GSE139242: Genes upregulated in infant CD4$^+$ thymocytes compared to CD4$^+$ T-cells in infant blood	2.24	1.10E-09	4.91E-10
GSE139242: Genes upregulated in infant CD4$^+$ thymocytes compared to CD4$^+$ T-cells in adult blood	2.31	1.10E-09	4.91E-10

Additionally, differential gene expression analyses revealed significantly increased expression of markers of immaturity and genes expressed during lymphocyte development, such as *CD34*, *KIT*, *DNTT*, and *CCR9* in canine CD4$^+$ PTCL compared to control canine CD4$^+$ lymphocytes (Table 6, Figure 21). Although *CD34* message is detected, protein expression of CD34 on these cells is consistently negative.

Table 6: Results of differential expression analysis of genes of immaturity between canine CD4$^+$ PTCL and control nodal CD4$^+$ T cells.

Cohort 1		
Gene	**Log2 fold change**	**Padj**
CD34	6.53	1.63E-09
KIT	3.26	0.002
DNTT	3.85	0.03
CCR9	5.27	6.67E-12
Cohort 2		
Gene	**Log2 fold change**	**Padj**
CD34	7.73	2.69E-14
KIT	3.04	0.0003
DNTT	3.06	0.08
CCR9	4.85	2.49E-11


A
Expression of Markers of Immaturity: Cohort 1
CCR9
CD34
DNTT
KIT
log10 Normalized Counts
Group
CD4_PTCL
CD4_CTRL
CD8_PTCL
CD8_CTRL
DN_PTCL
ns
***
**
1e+05
1e+03
1e+01
10000
1000
100
10
1

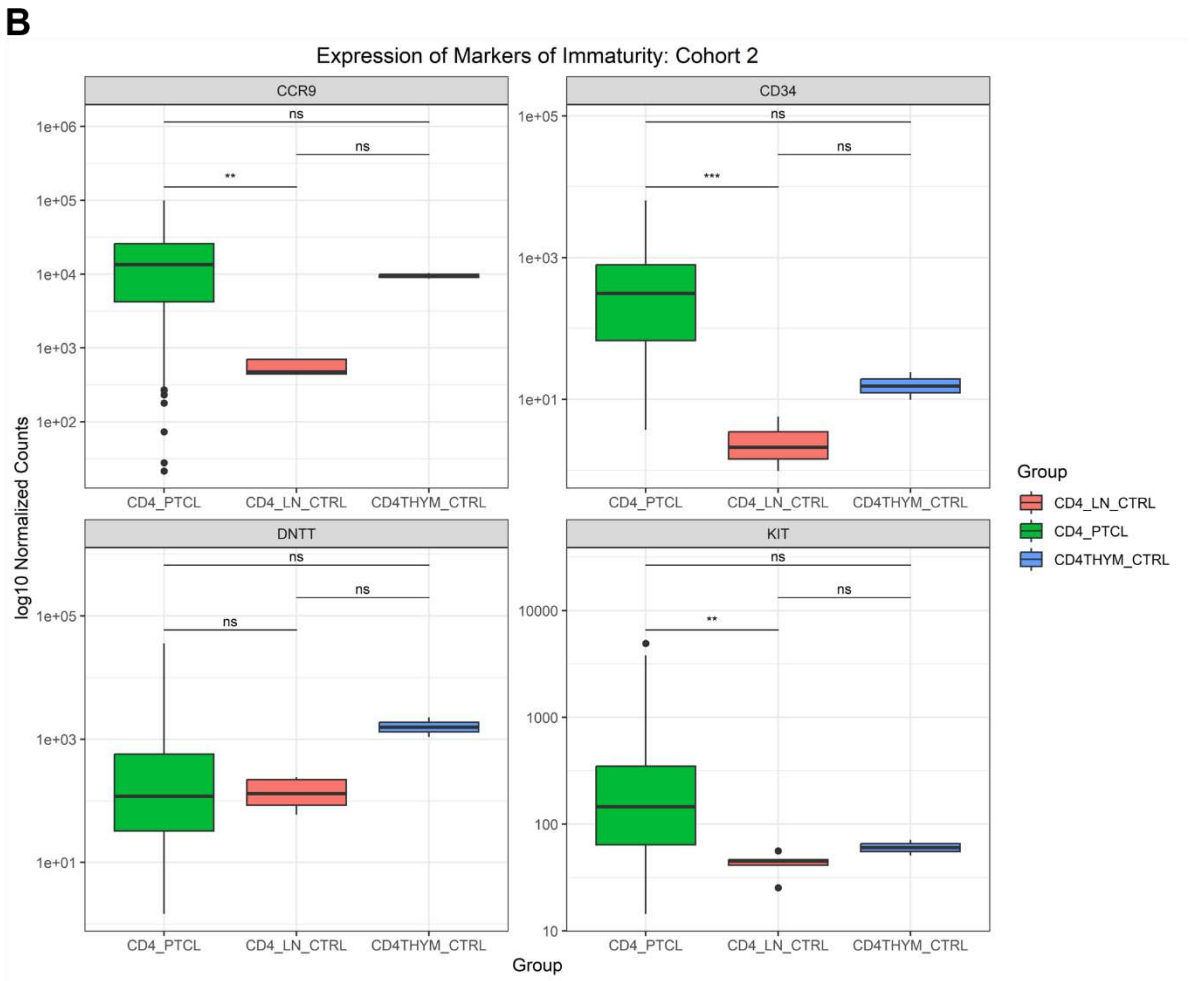


Figure 21: **Expression of genes of immaturity in CD4⁺ PTCL.** (**A**) A Wilcoxon signed-rank test comparing the means of normalized count data between canine PTCL subtypes and control nodal CD4⁺ and CD8⁺ lymphocytes in Cohort 1 (**A**) and between canine CD4⁺ PTCL and control nodal CD4⁺ lymphocytes and CD4⁺ thymocytes in Cohort 2 (**B**) identified significantly increased expression of *CCR9, CD34,* and *KIT* in CD4⁺ PTCLs. By differential expression analysis with DESeq2, *DNTT* was also significantly overexpressed in CD4⁺ PTCLs belonging to Cohort 1.

GSEA comparing the gene expression profile of canine CD4⁺ PTCL to gene sets associated with various mature murine T-cell lineages and naïve CD4⁺ T-cells (GSE14308) found no significant enrichment for Th1, Th2, Th17, or Treg signatures compared to naïve CD4⁺ T-cells in canine CD4⁺ PTCL (data for Cohort 1 available in Supplementary Table 3A and 4A of https://doi.org/10.1186/s12885-023-11762-w). Of note, there was one contradictory gene expression study (GSE22886) that showed negative enrichment of canine CD4⁺ PTCL for genes upregulated in naïve CD4⁺ T-cells compared to activated Th1 and Th2 cells.

Enrichment of individual canine CD4+ PTCL samples and control canine CD4+ lymphocyte samples for the expression profiles of various stages of human thymocyte development (GSE1460) and murine thymocyte development (GSE22601, GSE22886), resting and activated human immune cells (GSE22886), murine T-helper cell subsets (GSE14308), and the top 250 upregulated and downregulated genes in human CD4 SP infant thymocytes compared to CD4+ T-cells in adult and infant blood (from Helgeland et al.[15]) was evaluated by GSVA. Control CD4+ nodal T cells from Cohort 1 and 2 were negatively enriched for early T-cell progenitor and thymocyte gene signatures and strongly enriched for naïve T-cell signatures (Figure 22). Interestingly, control CD4+ thymocytes from Cohort 2 exhibited a similar enrichment profile to control CD4+ nodal lymphocytes, with the exception of being enriched for a set of genes upregulated in CD4+ thymocytes compared to circulating naïve CD4+ T cells, which the control CD4+ nodal lymphocytes were all negatively enriched for (Figure 22B). Two distinct clusters were identified within the CD4+ PTCL phenotype of Cohort 1 by unsupervised hierarchical clustering of GSVA scores (Figure 22A). The first group was characterized by strong enrichment of tumor cells for early T lymphocyte gene signatures, intrathymic T progenitor cells, and genes upregulated in DP thymocytes compared to naïve circulating CD4+ T-cells. Members of this group were variably enriched for CD4 SP thymocyte gene signatures and often either negatively enriched or non-enriched for genes upregulated in naïve CD4+ T-cells compared to activated Th1 and Th2 cells. Members of the second group observed by hierarchical clustering of GSVA within the CD4+ PTCL phenotype were largely negatively enriched or non-enriched for early T lymphocyte gene signatures, intrathymic T progenitor cells, and genes upregulated in DP thymocytes compared to naïve circulating CD4+ T-cells. Instead, this group showed enrichment for CD4 SP thymocyte gene signatures and genes upregulated in naïve CD4+ T-cells compared to activated Th1 and Th2 cells. In Cohort 2, CD4+ PTCLs were heterogeneously enriched for gene signatures of human early thymic progenitor, DP thymocyte, DN thymocyte, and early SP thymocyte gene signatures that tend to be negatively enriched in control nodal CD4+

lymphocytes and control CD4⁺ SP thymocytes, while a smaller subset of 4 CD4⁺ PTCLs were enriched for gene signatures of naïve T cells and TCR signaling signatures that also tended to be enriched in controls (Figure 22B). This variation within our canine CD4⁺ PTCL cases may suggest the existence of subtypes of canine CD4⁺ PTCL with varying degrees of T-cell differentiation, which comparisons to canine single-cell RNA-seq data may help further elucidate.

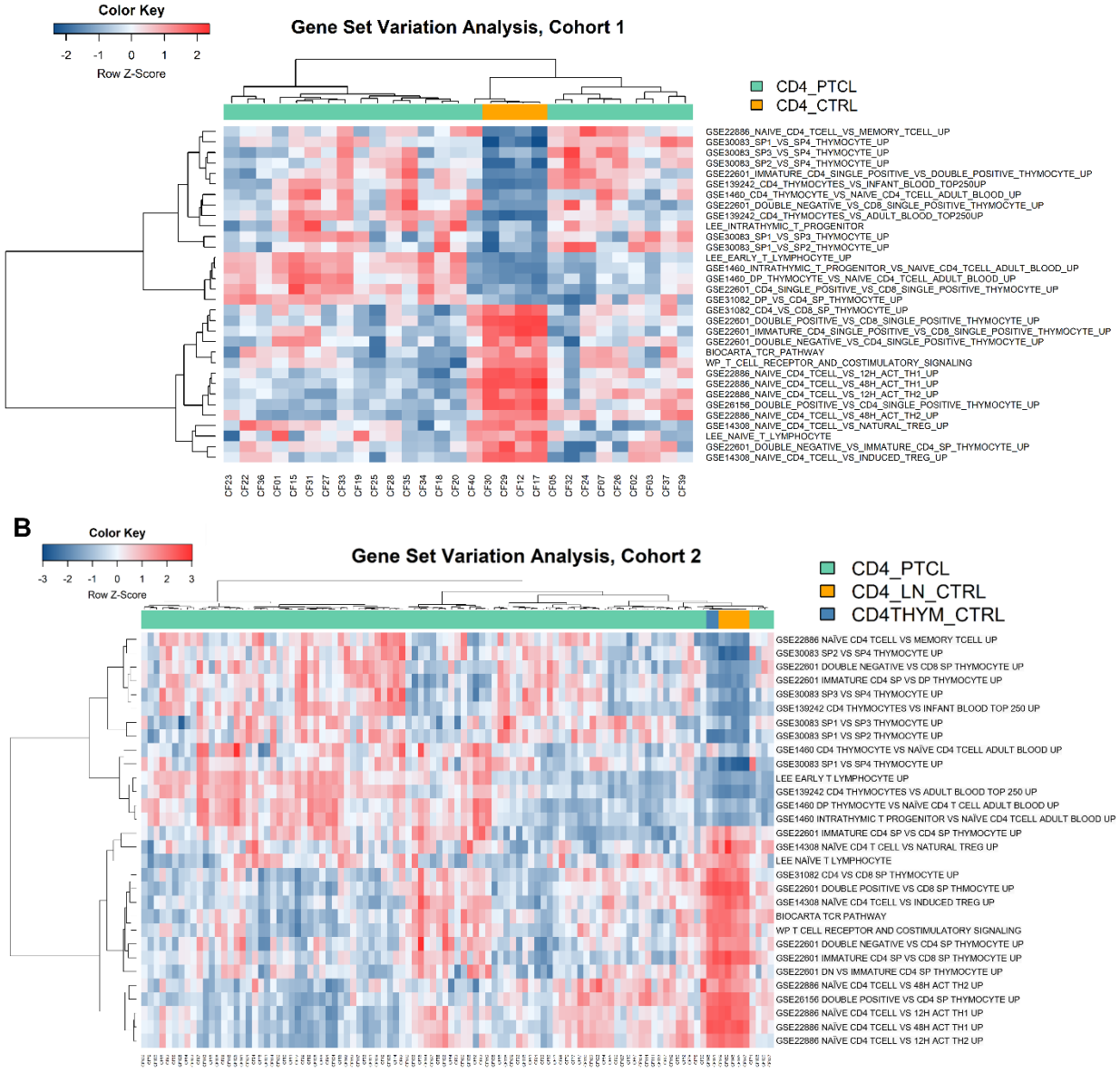


Figure 22: Gene set variation analysis (GSVA) identifies subsets of CD4⁺ PTCL with varying degrees of enrichment for human and murine T-cell differentiation stages. (**A**) Two distinct clusters were identified within the CD4⁺ PTCL phenotype of Cohort 1 by unsupervised hierarchical clustering of GSVA scores. One cluster was enriched for early T lymphocyte gene signatures, intrathymic T progenitor cells, and genes upregulated in DP thymocytes compared to naïve circulating CD4⁺ T-cells. The other cluster was enriched for SP thymocyte gene signatures and genes upregulated in

naïve CD4$^+$ T-cells compared to activated Th1 and Th2 cells. Control CD4$^+$ nodal lymphocytes were consistently enriched for naïve and mature T-cell gene signatures and negatively enriched for thymocyte and early T-cell progenitor gene signatures. (**B**) In Cohort 2, canine CD4$^+$ PTCLs are heterogeneously enriched for early thymic progenitor, DP thymocyte, DN thymocyte, and SP thymocyte gene signatures that tend to be negatively enriched in control nodal CD4$^+$ lymphocytes and CD4$^+$ SP thymocytes. A subset of 4 CD4$^+$ PTCLs also exhibits enrichment for some of the more mature gene signatures that are likewise enriched in controls.

Pseudotemporal trajectory analysis of a single-cell transcriptomic atlas identifies key genes involved in canine thymocyte and T-cell development

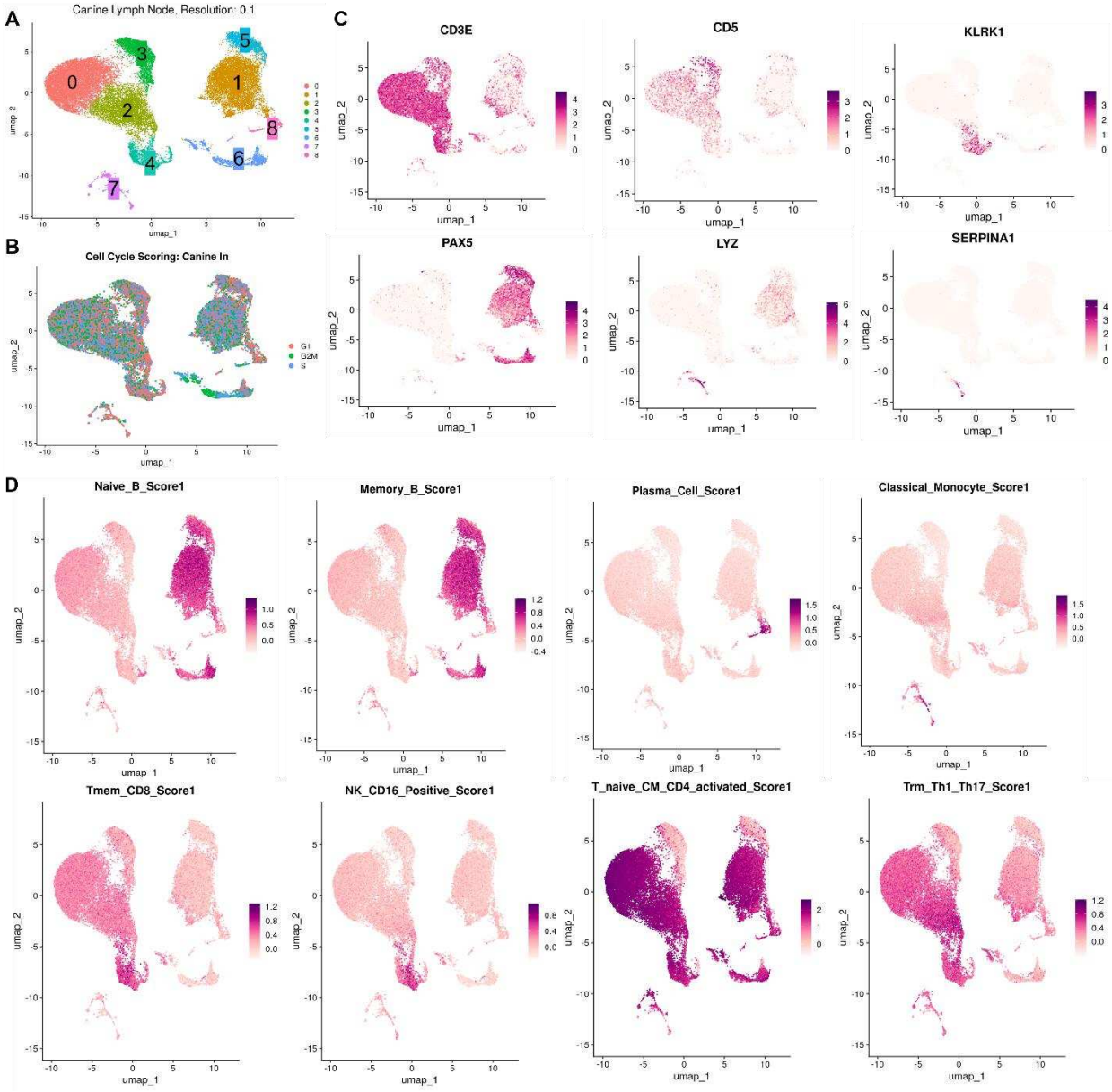
To compare CD4$^+$ PTCL to normal canine T cells at various stages of maturation similar to the analyses performed with human and murine data above, we first needed to generate an atlas of normal canine T-cell development. To this end, we utilized single-cell RNA-sequencing data of normal canine thymus and lymph node generated by another member of our lab, Dr. Adam Harris, which he collected from healthy 1-year-old dogs. Quality parameters for single-cell RNA-seq of normal canine thymus and lymph node are summarized in Supplemental Figure 8 and Supplemental Table 14. Single-cell RNA-seq data was collected from a median of 7,815 cells (range: 6,480-9,130) per sample of thymus and a median of 10,082 cells (range: 5,844-13,126) per sample of lymph node. The average number of genes detected per cell was 1,839 for thymus cells (range: 100-7,309) and 1,825 for lymph node cells (range: 102-7,876). Levels of mitochondrial gene expression per cell are a metric used to identify low quality of dying cells, as these cells often demonstrate considerable mitochondrial contamination. Mitochondrial genes comprised an average of 5% of all thymus cells (range: 0-96%) and an average of 6% of all lymph node cells (range: 0-93%). After filtering low quality cells, a median of 7,551 cells (range: 6,264-8,886) per sample of thymus and a median of 9,812 cells (range: 5,576-12,823) per sample of lymph node remained, with an average number of 1,854 genes detected per cell for thymus cells (range: 302-5,990) and 1,840 for lymph node cells (range: 303-5,991). After filtering, mitochondrial genes comprised an average of 4% of all thymus cells and an average of 5% of all lymph node cells.

Nine distinct clusters were identified in single-cell RNA-seq data from normal canine lymph node samples (Figure 23A). Cell cycle scoring identified a high number of cycling cells in cluster 6 (Figure 23B). Differential expression of markers of interest identified cells expressing T-cell markers *CD3* and *CD5* (clusters 0, 2, 3, and 4), cells expressing CD8+ T-cell/NK-cell marker *KLRK1* (cluster 4), cells expressing B-cell marker *PAX5* (clusters 5, 1, and 6), and cells expressing myeloid markers *LYZ* and *SERPINA1* (cluster 7) (Figure 23C). The full list of genes evaluated is available on Rpubs: https://rpubs.com/eileenowens/LNAnnot1.

Module scoring, in which the relative expression of a feature set is calculated for each cluster to identify clusters enriched for those features, was performed using gene sets derived from a human single-cell immune cell atlas (Conde et al.[22]). Cells enriched for naïve and memory B-cell genes were identified in clusters 5, 1, and 6; plasma cell genes in cluster 8; monocyte genes in cluster 7; memory T cell and CD8+ T cell genes in clusters 0, 2, and especially 4; NK cell genes in cluster 4; and genes of tissue-resident memory, Th1, and Th17 cells predominantly in clusters 0, 2, and 4. Genes associated with naïve, activated, and memory $CD4^+$ T cells in this human atlas were more ubiquitously expressed in our canine single-cell data (Figure 23D). Full module scoring results are available on Rpubs: https://rpubs.com/eileenowens/LNAnnot3.

GSEA using gene sets derived from a human single-cell immune cell atlas (Conde et al.[22]), a single-cell atlas of circulating canine leukocytes (Ammons et al.[23]), and gene sets of genes upregulated by various naïve and activated T-cell subsets from the Broad Institute Molecular Signatures Database[14] (Supplemental Table 12) revealed enrichment of clusters 0, 2, and 3 for assorted T-cell subtypes, enrichment of clusters 1 and 5 for various B-cell subtypes, and enrichment of cluster 4 for CD8 T and NK cell gene sets. No gene sets were significantly enriched in clusters 6, 7, and 8. Based on these annotations, clusters 0, 2, 3, and 4 were ultimately subset from this atlas

and combined with single-cell data from canine thymus samples for trajectory analysis of the T-cell lineage.



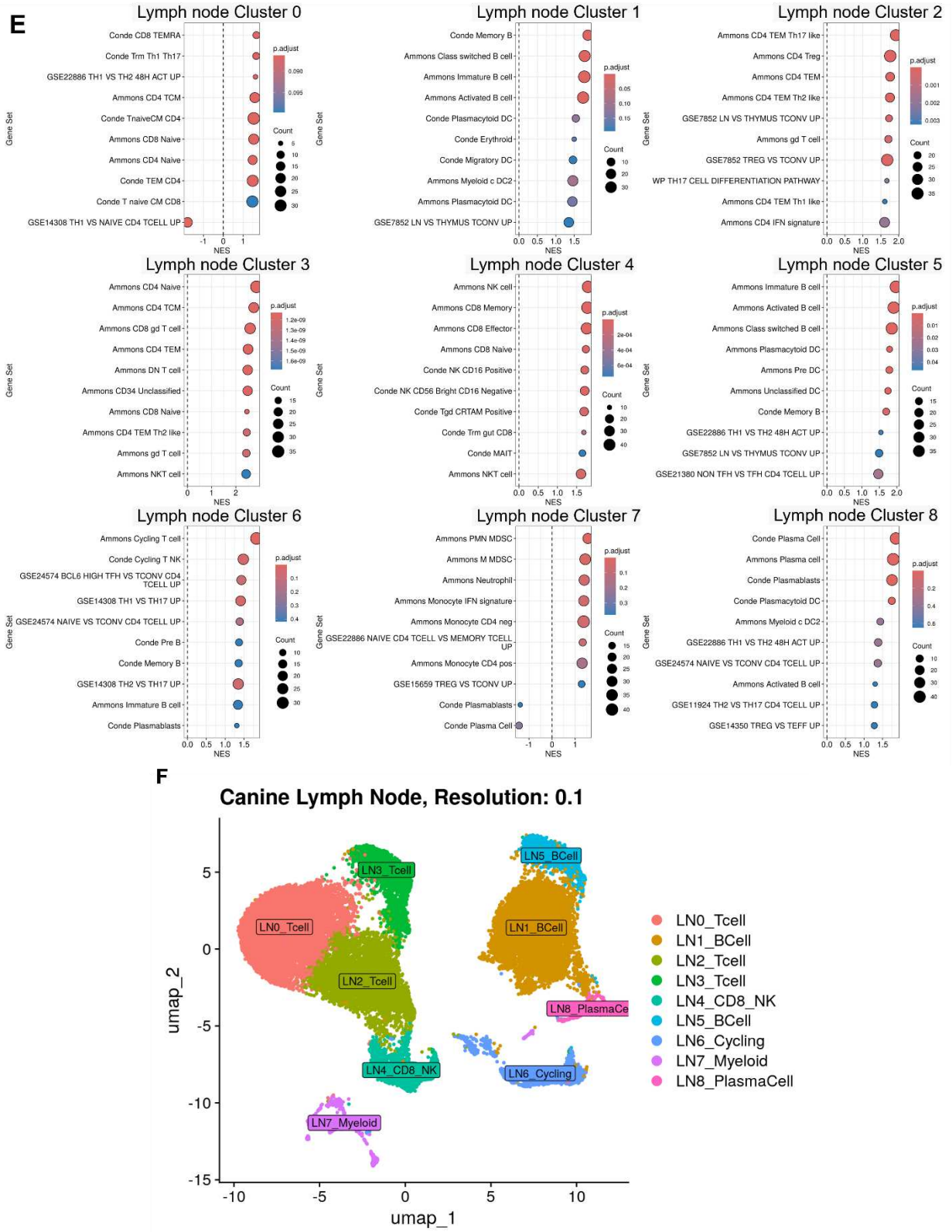


Figure 23: Annotation of single-cell RNA-seq clusters from normal canine lymph node samples. (**A**) Clustering performed with Seurat at 0.1 resolution with n.neighbors = 30 and min.dist = 0.3 identified 9 distinct clusters in the normal canine lymph node. (**B**) Cell cycle scoring identified a high number of cycling cells in cluster 6. (**C**) Differential

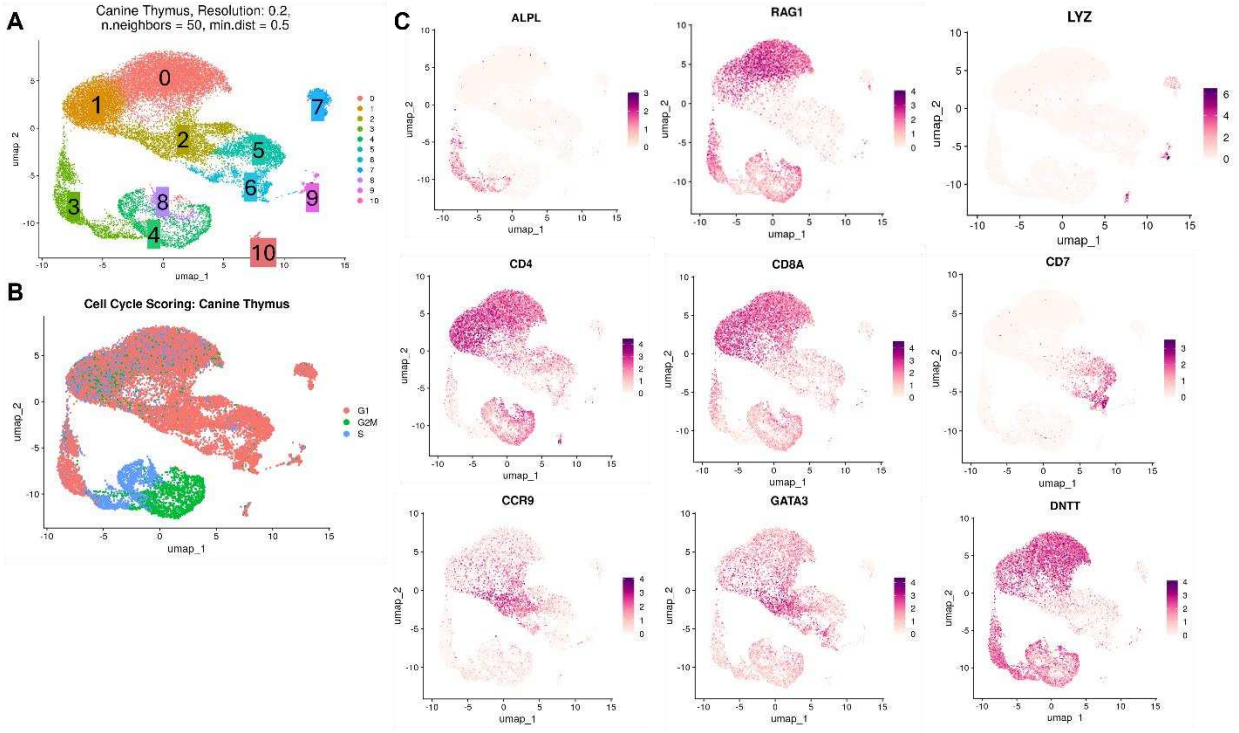
expression of markers of interest identified cells expressing T-cell markers *CD3* and *CD5* (clusters 0, 2, 3, and 4), cells expressing $CD8^+$ T-cell/NK-cell marker *KLRK1* (cluster 4), cells expressing B-cell marker *PAX5* (clusters 5, 1, and 6), and cells expressing myeloid markers *LYZ* and *SERPINA1* (cluster 7). (**D**) Module scoring using gene signatures derived from a human immune cell atlas (Conde et al.). (**E**) GSEA using gene sets derived from a human single-cell immune cell atlas (Conde et al.), a single-cell atlas of circulating canine leukocytes (Ammons et al.), and gene sets of genes upregulated by various naïve and activated T-cell subsets from the Broad Institute Molecular Signatures Database. Each enrichment plot represents the gene sets from these references that are positively or negatively enriched in each lymph node cluster, with cluster numbers corresponding to those in (A). (**F**) Final annotations assigned to single-cell clusters of normal canine lymph node samples prior to subsetting T cell clusters and merging with canine thymus single-cell data.

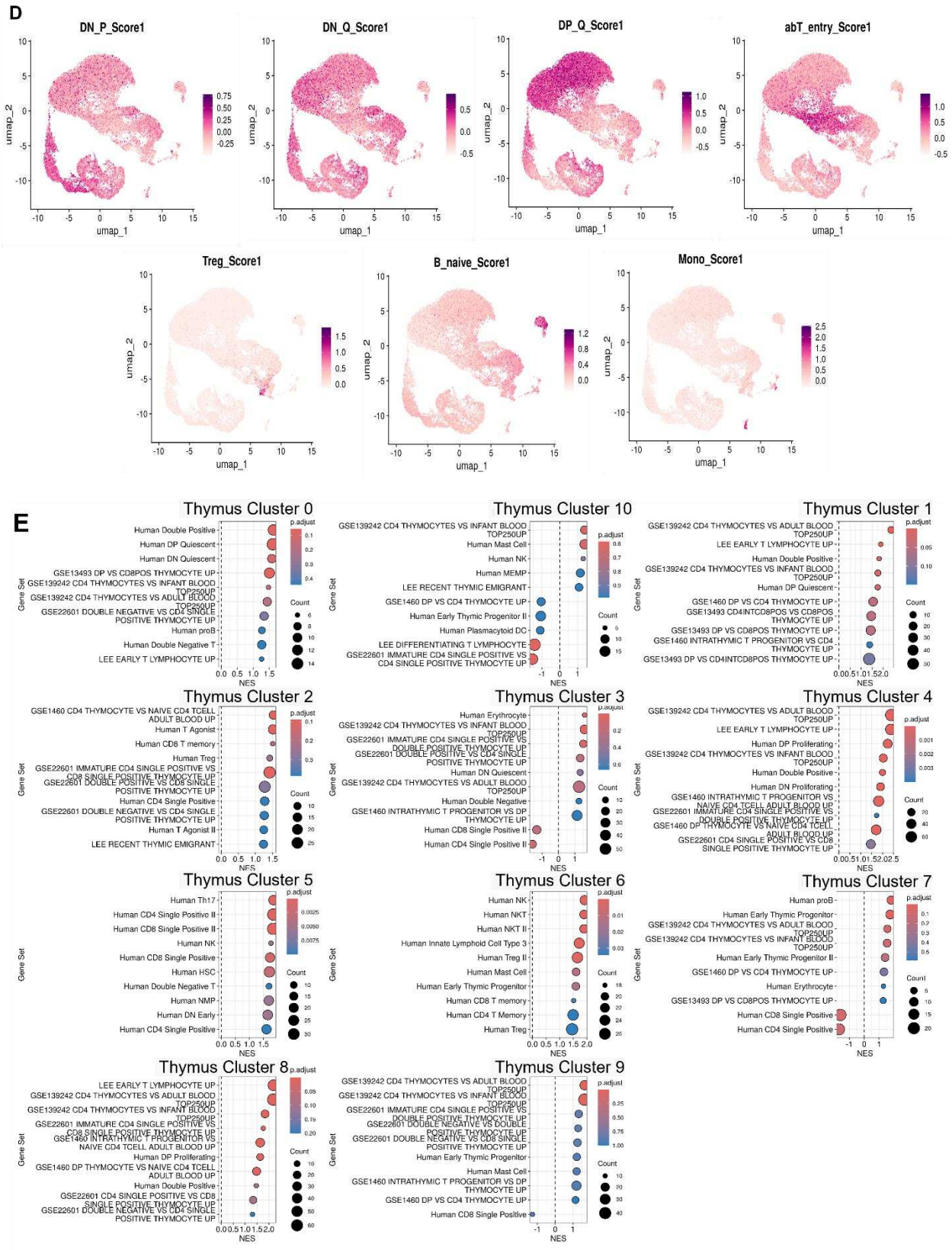
Eleven distinct clusters were identified in single-cell RNA-seq data from normal canine thymus samples (Figure 24A). Cell cycle scoring identified a high number of cycling cells in clusters 4 and 8 (Figure 24B). Differential expression of markers of interest identified cells expressing hematopoietic stem cell marker *ALPL* in cluster 3; cells expressing TCR rearrangement gene *RAG1* in clusters 0, 1, 3, 4, and 8; cells expressing myeloid marker *LYZ* in clusters 8 and 10; a large proportion of *CD4* and *CD8* expressing cells in clusters 0 and 1; cells expressing more mature T-cell marker *CD7* in clusters 5 and 6; cells expressing *CCR9*, a marker of the DP-to-SP thymocyte transition,[21] in cluster 2; relatively ubiquitous expression of transcription factor *GATA3* that particularly concentrated in cluster 2; and expression of marker of immaturity *DNTT* in clusters 0, 1, 3, 4, and 8 (Figure 24C). The full list of genes evaluated is available on Rpubs: https://rpubs.com/eileenowens/thymusAnnot1.

Module scoring using gene signatures derived from a human single-cell thymus atlas (Park et al.[21]) identified cells enriched for proliferating DN thymocyte genes predominantly in cluster 3, quiescent DP thymocyte genes predominantly in clusters 0 and 1, the $CCR9^{high}$ Tαβ (entry) stage between late DP and SP thymocytes in cluster 2, Treg genes in cluster 6, naïve B cell genes in cluster 7, monocyte genes in clusters 9 and 10, and relatively ubiquitous expression of genes associated with quiescent DN thymocytes. Full module scoring results are available on Rpubs: https://rpubs.com/eileenowens/ThymAnnot3.

GSEA using gene sets derived from a human single-cell thymus atlas (Park et al.[21]) and a curated list of gene sets containing genes upregulated by various T-cell progenitor and thymocyte

subsets from the Broad Institute Molecular Signatures Database[14] (Supplemental Table 12) revealed significant enrichment of cluster 1 for early T lymphocyte and DP thymocyte gene sets; cluster 4 for early T lymphocyte, intrathymic T progenitor, and proliferating DN and DP gene sets; cluster 5 for early and late CD4 and CD8 SP cells and gene sets of more mature lymphoid cells like Th17 cells and NK cells; cluster 6 for primarily NK, NKT, Treg, and memory T cell gene sets; and cluster 8 for early T lymphocyte, CD4 lineage, proliferating DP, and intrathymic T progenitor cell gene sets. No gene sets were significantly enriched in clusters 0, 2, 3, 7, 9, or 10.




D
DN_P_Score1
DN_Q_Score1
DP_Q_Score1
abT_entry_Score1
Treg_Score1
B_naive_Score1
Mono_Score1
umap_1
umap_2
E
Thymus Cluster 0
Thymus Cluster 10
Thymus Cluster 1
Thymus Cluster 2
Thymus Cluster 3
Thymus Cluster 4
Thymus Cluster 5
Thymus Cluster 6
Thymus Cluster 7
Thymus Cluster 8
Thymus Cluster 9
Gene Set
NES
p.adjust
Count

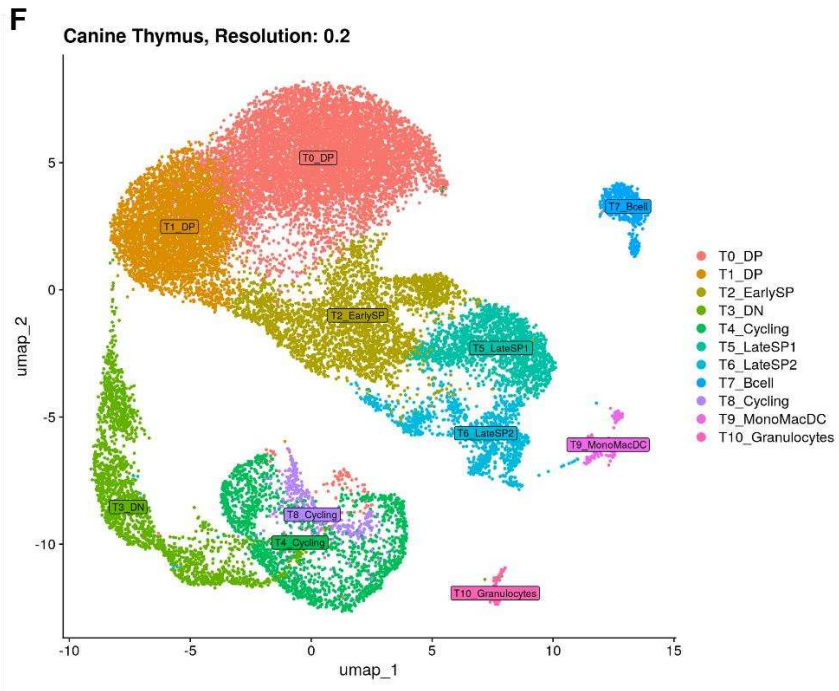


Figure 24: Annotation of single-cell RNA-seq clusters from normal canine thymus samples. (**A**) Clustering performed with Seurat at 0.2 resolution with n.neighbors = 50 and min.dist = 0.5 identified 11 distinct clusters in the normal canine thymus. (**B**) Cell cycle scoring identified a high number of cycling cells in clusters 4 and 8. (**C**) Differential expression of markers of interest identified cells expressing hematopoietic stem cell marker *CD34* in cluster 3; cells expressing TCR rearrangement gene *RAG1* in clusters 0, 1, 3, 4, and 8; cells expressing myeloid marker *LYZ* in clusters 8 and 10; a large proportion of *CD4* and *CD8* expressing cells in clusters 0 and 1; cells expressing more mature T-cell marker *CD7* in clusters 5 and 6; cells expressing *CCR9,* a marker of the DP-to-SP transition, in cluster 2; relatively ubiquitous expression of transcription factor *GATA3* that particularly concentrates in cluster 2; and expression of marker of immaturity *DNTT* in clusters 0, 1, 3, 4, and 8. (**D**) Module scoring using gene signatures derived from a human single-cell thymus atlas (Park et al.). (**E**) GSEA using gene sets derived from a human single-cell thymus atlas (Park et al.) and gene sets of genes upregulated by various T-cell progenitor and thymocyte subsets from the Broad Institute Molecular Signatures Database. Each enrichment plot represents the gene sets from these references that are positively or negatively enriched in each lymph node cluster, with cluster numbers corresponding to those in (A). (**F**) Final annotations assigned to single-cell clusters of normal canine thymus samples prior to subsetting T cell clusters and merging with canine lymph node single-cell data.

To prepare for pseudotime trajectory analysis of the T-cell lineage, only T lineage-associated clusters from the individual canine thymus and lymph node atlases were isolated and merged into one new atlas. These included clusters “T0_DP”, “T1_DP”, “T2_EarlySP”, “T4_Cycling”, “T5_LateSP1”, and “T6_LateSP2” from the thymus atlas and clusters “LN0_Tcell”, “LN2_Tcell”, “LN3_Tcell”, and “LN4_CD8_NK” from the lymph node atlas. Normalization, dimensionality reduction, and clustering of this merged atlas produced eleven distinct clusters (Figure 25A). In general, cells that clustered together in the individual thymus and lymph node atlases remained

clustered in the merged atlas, with the exception of thymus cluster 10, which was split into two distinct clusters, 9 and 10, in the merged atlas (Figure 25B).

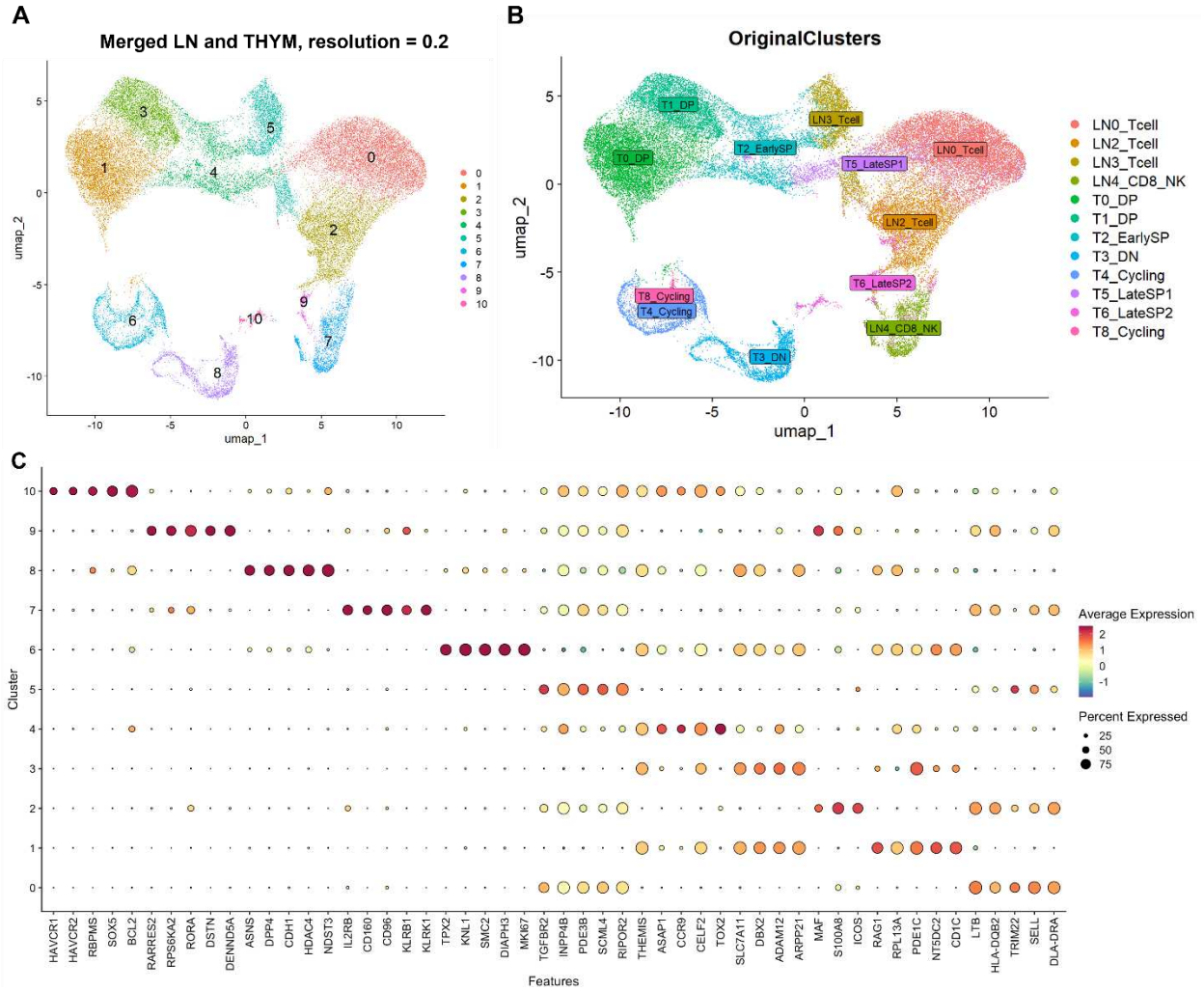


Figure 25: Annotation of single-cell RNA-seq clusters of merged normal canine thymocyte and nodal T cell samples. (**A**) Clustering performed with Seurat at 0.2 resolution with n.neighbors = 30 and min.dist = 0.3 identified 11 distinct clusters in the merged populations of T-cell progenitors, thymocytes, and T cells from normal canine thymus and lymph node. (**B**) Previous annotations assigned to the thymus and lymph node overlaid onto the UMAP of merged thymus and lymph node data sets reveal close associations between SP thymocyte clusters and naïve nodal T-cell clusters, and SP thymocyte clusters and mature, activated CD8$^+$ and NK cell clusters. (**C**) Top marker expression by cluster, as determined by filtering results of Seurat::FindAllMarkers() to prioritize the top 5 unique features for each cluster. The percentage of cells in a cluster expressing the gene is denoted by dot size, while the average expression level of the gene is indicated by color. Orange and red dots represent high expression of that gene compared to all other clusters, yellow dots indicate unchanged expression compared to all other clusters, and green and blue dots indicate decreased expression compared to all other clusters.

Cluster 0 of our combined atlas, originally derived from the lymph node, was characterized by high expression of T-cell homing molecules *SELL* (L-selectin) and *CCR7*, which are typically most expressed in naïve T cells (Figure 26A).[27] It was also most enriched for naïve CD4$^+$ and CD8$^+$ T-cell

gene signatures derived from a human single-cell immune cell atlas (Conde et al.[22]) and a single-cell atlas of canine circulating leukocytes (Ammons et al.[23]) (Table 7, Figure 26B). Interestingly, when compared to gene signatures derived from a single-cell atlas of human thymus (Park et al.[21]), it demonstrated significant enrichment for hematopoietic stem cell, late SP thymocyte, and early DN cell signatures (Figure 26C). Leading edge analysis of these enrichment results identified *HOPX* as the top gene contributing to its enrichment for hematopoietic stem cells. *HOPX* marks naïve T cells destined for subsequent effector differentiation and plays a role in regulating Tregs induced in the periphery. It is also upregulated early in development as a regulator of stem cell proliferation and differentiation, which likely explains the crossover observed here. *LTB* is generally regarded as a marker of T-cell activation,[28] but notably its expression can be observed in late SP thymocytes and naïve T cells in human single-cell atlases of thymus and cross-tissue immune cells.[21,22] *LY6E*, which plays a role in T-cell activation,[29] was also highly expressed by this cluster, which could indicate a state of transition from naïve to activated T cells for this cluster.

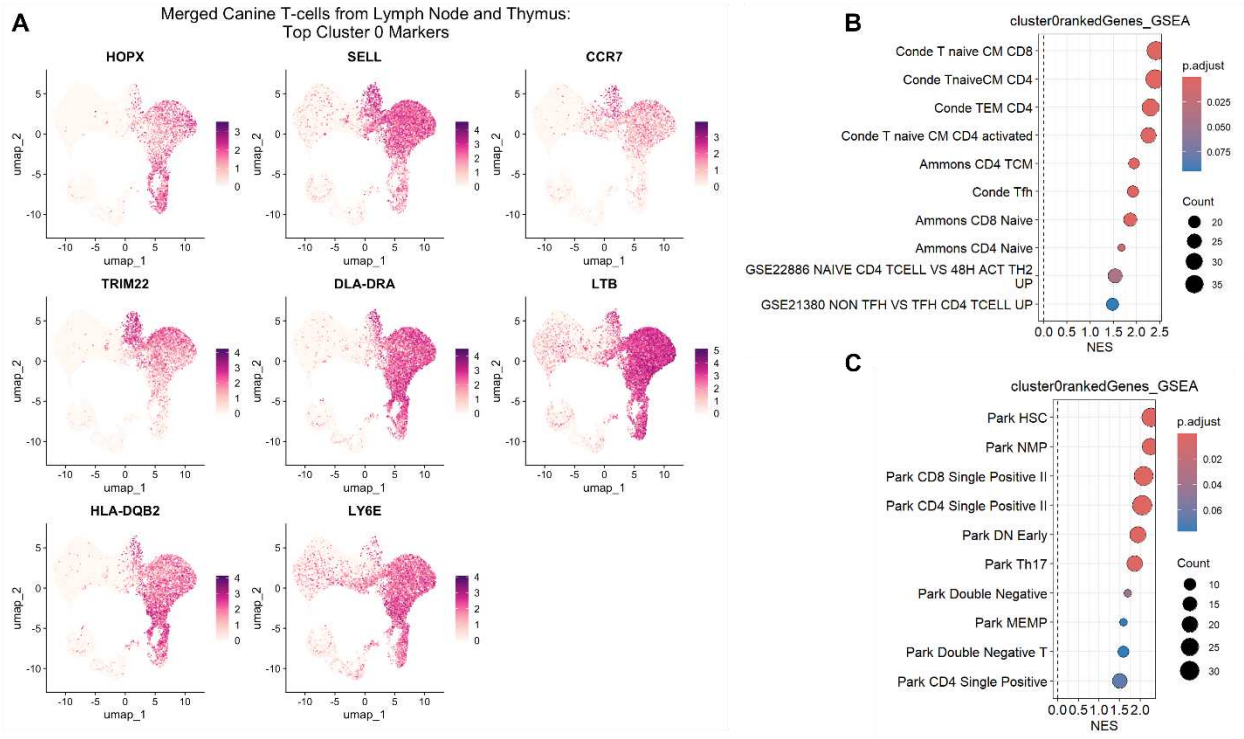


Figure 26: Annotation details of cluster 0. (**A**) Expression of key developmental markers and top differentially expressed markers in cluster 0. (**B-C**) GSEA enrichment results for gene sets derived from a human immune cell atlas (Conde et al.), a canine circulating leukocyte atlas (Ammons et al.), and gene sets of naïve and activated T-cell subset gene signatures from the Broad Institute Molecular Signatures Database (**B**) and derived from a human single-cell thymus atlas (Park et al.) and sets of genes upregulated by various T-cell progenitor and thymocyte subsets from the Broad Institute Molecular Signatures Database (**C**).

Cluster 1 of our combined atlas, originally derived from the thymus, was characterized by high expression of both *CD4* and *CD8A* consistent with DP thymocytes (Figure 27A). They also highly expressed *DNTT*, which plays a role in TCR rearrangement and is normally expressed in DN and DP thymocytes before being downregulated as thymocytes mature to SP cells.[30] *CD1C* was also highly expressed in this cluster, and in macaques, CD1 glycoprotein expression has been described as a discriminatory marker of thymocyte development, appearing in thymocytes prior to β-selection and becoming progressively downregulated as thymocytes continue to mature.[31] *NT5DC2*, another one of the top genes upregulated by this cluster, is associated with cell proliferation,[32] consistent with the proliferative burst undergone by DP thymocytes. Cells in this cluster also expressed *THEMIS*, an important regulator of positive selection and lineage commitment that is upregulated in late DN3 thymocytes and remains highly expressed in pre-selection DP thymocytes before being

downregulated after cells proceed through positive selection and remaining downregulated in mature SP cells.[33] By GSEA, cluster 1 was most enriched for gene sets associated with DP thymocytes (Table 7, Figure 27B-C).

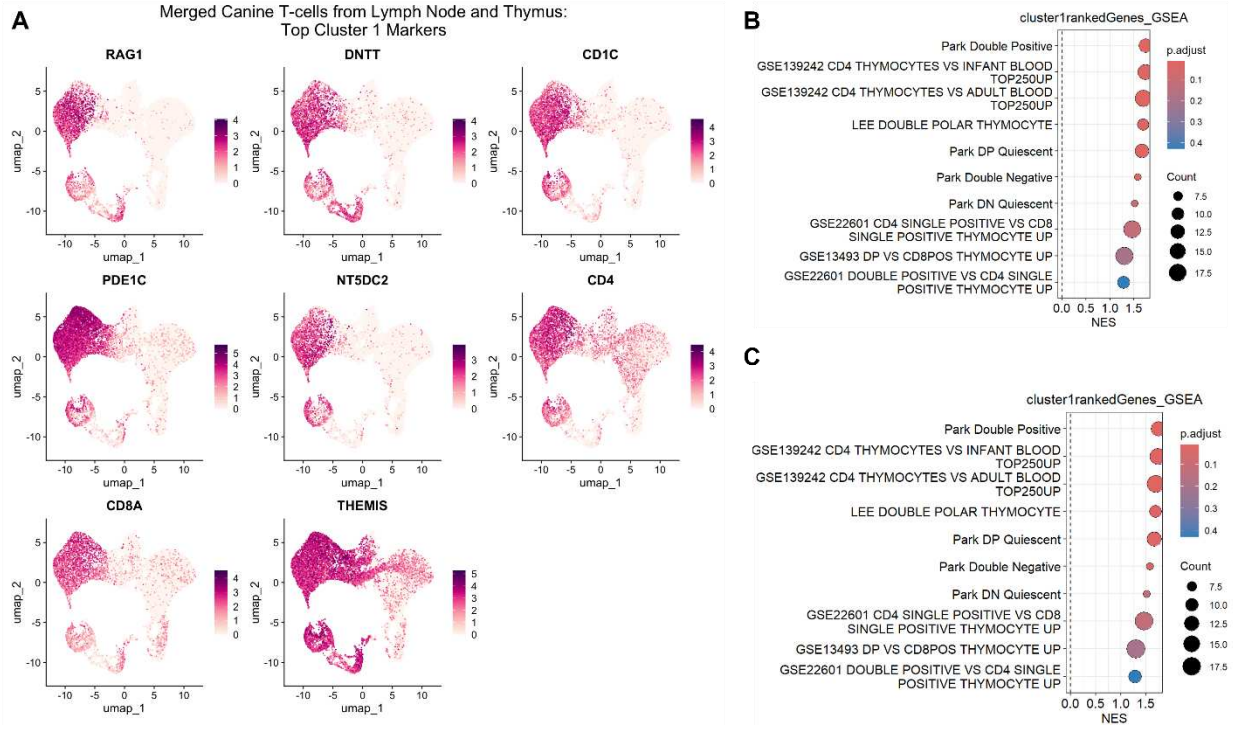


Figure 27: Annotation details of cluster 1. (**A**) Expression of key developmental markers and top differentially expressed markers in cluster 1. (**B**) No gene sets derived from a human immune cell atlas (Conde et al.), a canine circulating leukocyte atlas (Ammons et al.), and gene sets of naïve and activated T-cell subset gene signatures from the Broad Institute Molecular Signatures Database were significantly enriched in cluster 1. (**C**) GSEA enrichment results for gene sets derived from a human single-cell thymus atlas (Park et al.) and sets of genes upregulated by various T-cell progenitor and thymocyte subsets from the Broad Institute Molecular Signatures Database. “Park Double Positive”, “GSE139242 CD4 THYMOCYTES VS NFANT BLOOD TOP250UP”, “GSE139242 CD4 THYMOCYTES VS ADULT BLOOD TOP250UP”, “Park DP Quiescent”, and “LEE DOUBLE POLAR THYMOCYTE” were significantly enriched (padj < 0.05).

Cluster 2 of our combined atlas, originally derived from the lymph node, was characterized by expression of several markers associated with T-cell activation and maturity, and tended to be enriched for gene signatures associated with mature, activated, and fully differentiated T-cell subsets (Table 7, Figure 28). Notably, this cluster exhibited expression of T-cell deactivator *CTLA4*, supporting that cells in this cluster represent post-activated T cells. Expression of *MAF*, a transcription factor highly expressed by terminally differentiated T cells after TCR activation,[34–36] *IL2RA*, which in humans and dogs is associated with an activated lymphocyte phenotype,[37–39] and

ICOS, one of the costimulatory signals of T-cell activation,[40] also distinguished cells in this cluster from other lymph node-derived clusters in our atlas. This cluster also exhibited high expression of *S100A8*, a TLR4 agonist normally associated with myeloid cells[41,42] whose role in the T-cell lineage is unclear.

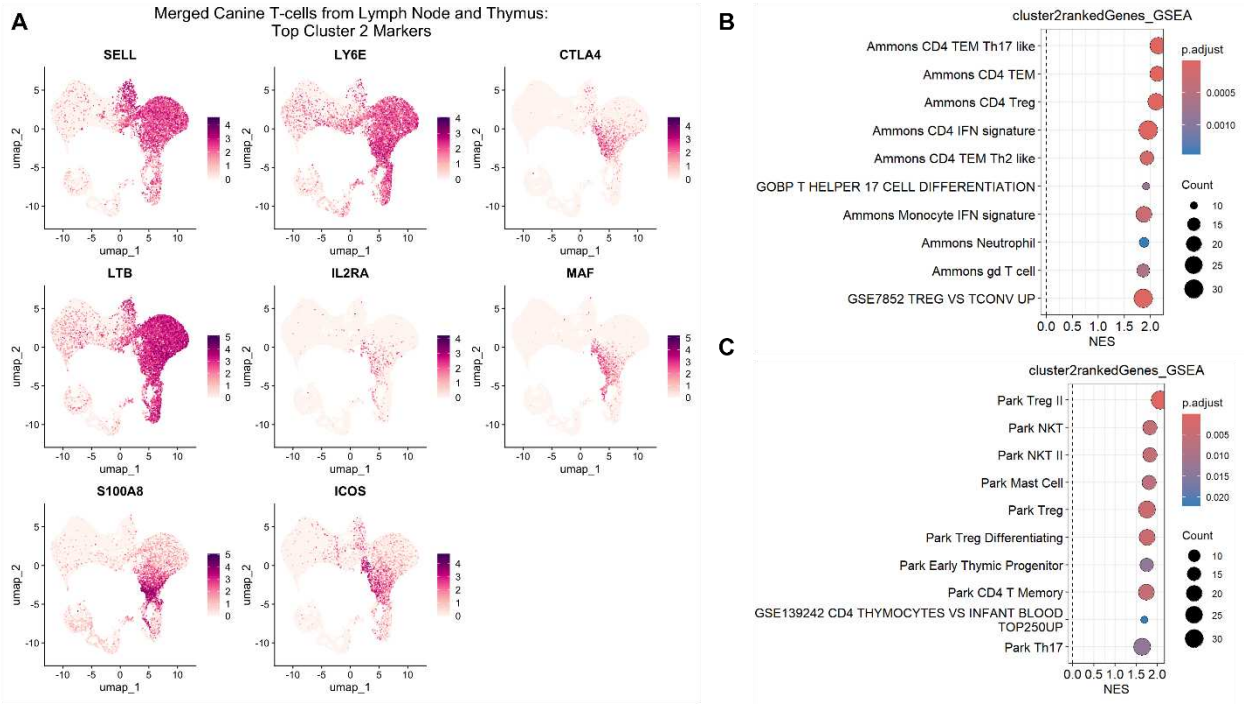


Figure 28: Annotation details of cluster 2. (**A**) Expression of key developmental markers and top differentially expressed markers in cluster 2. (**B-C**) GSEA enrichment results for gene sets derived from a human immune cell atlas (Conde et al.), a canine circulating leukocyte atlas (Ammons et al.), and gene sets of naïve and activated T-cell subset gene signatures from the Broad Institute Molecular Signatures Database (**B**) and derived from a human single-cell thymus atlas (Park et al.) and sets of genes upregulated by various T-cell progenitor and thymocyte subsets from the Broad Institute Molecular Signatures Database (**C**).

Cluster 3 of our combined atlas, originally derived from the thymus, was characterized by high expression of both *CD4* and *CD8A* consistent with DP thymocytes (Figure 29A) and was enriched for gene signatures of DP thymocytes, early T lymphocytes, and intrathymic progenitor cells by GSEA (Figure 29C). Many of the top markers expressed by this cluster were also highly expressed by DP thymocytes in cluster 1.

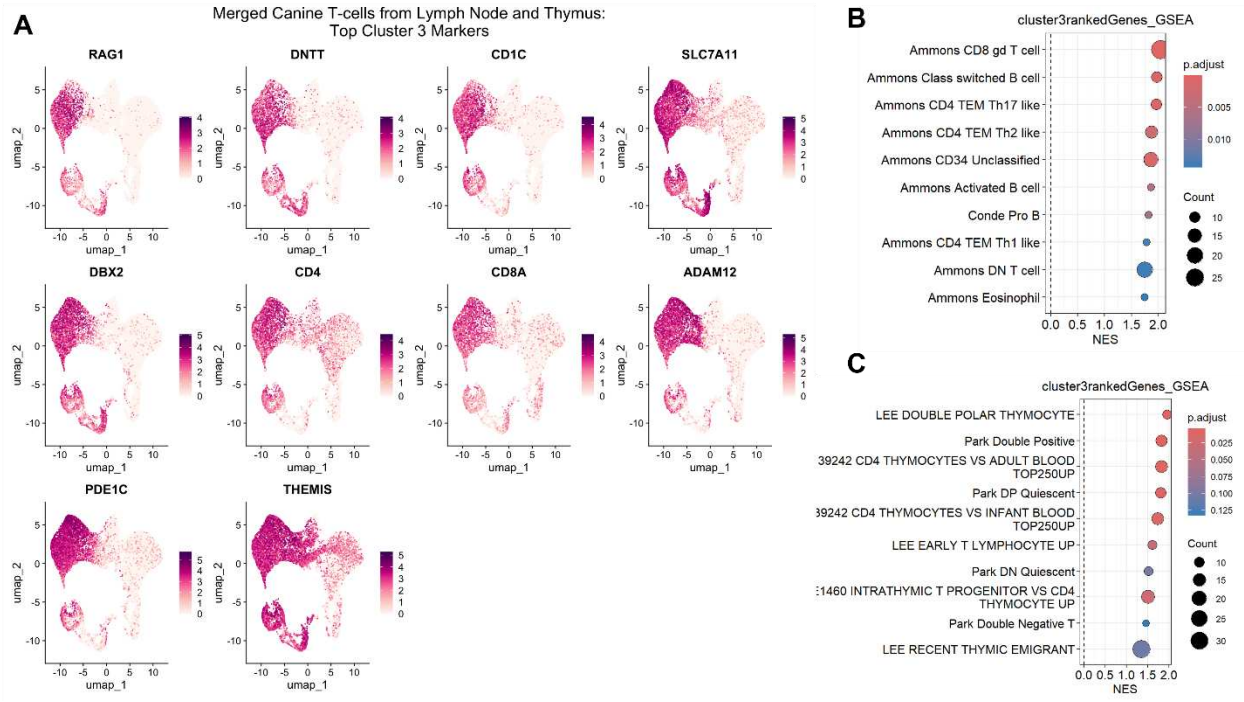


Figure 29: Annotation details of cluster 3. (**A**) Expression of key developmental markers and top differentially expressed markers in cluster 3. (**B-C**) GSEA enrichment results for gene sets derived from a human immune cell atlas (Conde et al.), a canine circulating leukocyte atlas (Ammons et al.), and gene sets of naïve and activated T-cell subset gene signatures from the Broad Institute Molecular Signatures Database (**B**) and derived from a human single-cell thymus atlas (Park et al.) and sets of genes upregulated by various T-cell progenitor and thymocyte subsets from the Broad Institute Molecular Signatures Database (**C**).

Cluster 4 of our combined atlas, originally derived from the thymus, was significantly enriched for CD4$^+$ SP thymocyte gene signatures by GSEA (Table 7, Figure 30C) and uniquely distinguished from other thymus clusters by its expression of *CCR9, TOX2,* and higher expression of *ASAP1* and *GATA3* (Figure 30A). In humans, *CCR9* is a marker of the DP to SP transition in the thymus, expressed by a distinctive CCR9^{high}, α/β TCR-expressing DP stage before their divergence into mature CD4$^+$ and CD8$^+$ SP cells.[21] *TOX2* is a member of the thymocyte-selection-associated high mobility group box (TOX) family composed of *TOX, TOX2, TOX3*, and *TOX4*. While best known for its role in promoting NK cell and follicular T-helper cell differentiation,[43] one study found that *TOX2* is also upregulated in DP thymocytes following positive selection.[44] Interestingly, *TOX2* is also involved in a fusion event in a subset of canine CD4$^+$ PTCL, as will be discussed more extensively in Chapter Three. Expression of transcription factor *GATA3* in late DP cells promotes development into

CD4+ SP T cells.[45,46] This cluster also highly expressed *CELF2*, a gene that in humans increases in expression during the DN to DP thymocyte transition and remains high in SP thymocytes.[47] The role of *ASAP1*—another gene highly expressed by this cluster—in thymocyte maturation is unclear. Taken together, these findings suggest that cluster 4 may represent late DP or early SP cells that have passed positive selection.

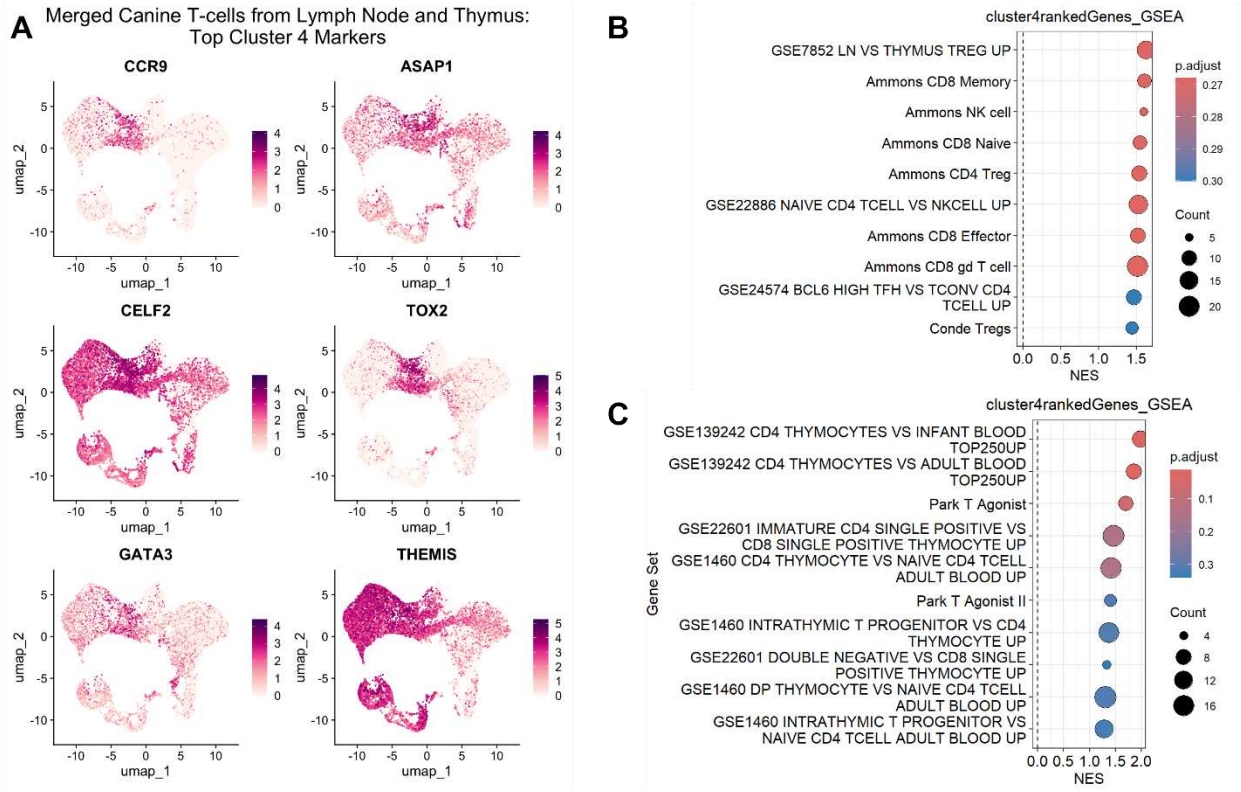


Figure 30: Annotation details of cluster 4. (**A**) Expression of key developmental markers and top differentially expressed markers in cluster 4. (**B**) No gene sets derived from a human immune cell atlas (Conde et al.), a canine circulating leukocyte atlas (Ammons et al.), and gene sets of naïve and activated T-cell subset gene signatures from the Broad Institute Molecular Signatures Database were significantly enriched in cluster 4. (**C**) GSEA enrichment results for gene sets derived from a human single-cell thymus atlas (Park et al.) and sets of genes upregulated by various T-cell progenitor and thymocyte subsets from the Broad Institute Molecular Signatures Database. "GSE139242 CD4 THYMOCYTES VS NFANT BLOOD TOP250UP" and "GSE139242 CD4 THYMOCYTES VS ADULT BLOOD TOP250UP" were significantly enriched (padj < 0.05).

Cluster 5 of our combined atlas, originally derived from the lymph node, was characterized by enrichment for naïve T-cell, SP thymocyte, recent thymic emigrant, Treg, and memory T-cell signatures (Table 7, Figure 31B-C) and expressed several shared markers with cluster 0, including *SELL, CCR7,* and *TRIM22* (Figure 31A). Leading edge analysis identified *ITGA1*, the gene encoding

integrin α1β1 (also known as VLA-1), as one of the top genes contributing to the enrichment of this cluster for memory T cell gene sets by GSEA. VLA-1 has been shown to only be expressed >6 days after T-cell activation and remains expressed on memory T cells.[48] Cluster 5 had the highest expression of *RIPOR2, PDE3B*, *TGFBR2,* and *SCML4* of any of the lymph node clusters in our atlas (Figure 25C, Figure 31A). *RIPOR2*, which encodes FAM65B, is highly expressed by naïve T cells and maintains them in a quiescent state, requiring downregulation before T cells can proliferate following engagement of their TCR by antigen.[49] *PDE3B* is expressed by naïve T cells and is downregulated during terminal differentiation to Treg cells through binding of transcription factor *FOXP3*.[50] TGF-β signaling is important in maintaining naïve T cell quiescence and is intact in these cells until it is overcome by strong TCR stimulation. *TGFBR2* transcript expression is downregulated in both human and murine $CD4^+$ T cells following activation through TCR signaling,[51] so the strong upregulation of *TGFBR2* in cluster 5 could support these cells existing in a naïve state. The role of *SMCL4* in T-cell differentiation is not well-characterized, although one study in mice found that this gene was upregulated in $CD8^+$ Trm cells, and its knockout hampered antitumor immunity.[52] Given the upregulation of several markers of naïve T cells, the close association of this cluster with our thymus-derived clusters, and the lack of expression of some of the T-cell activation genes appreciated in cluster 2, we ultimately annotated this group as naïve T cells. However, we acknowledge that this cluster also expressed some memory T cell-associated genes, and so this annotation may not fully capture the biologic cell type comprising this cluster of our atlas.

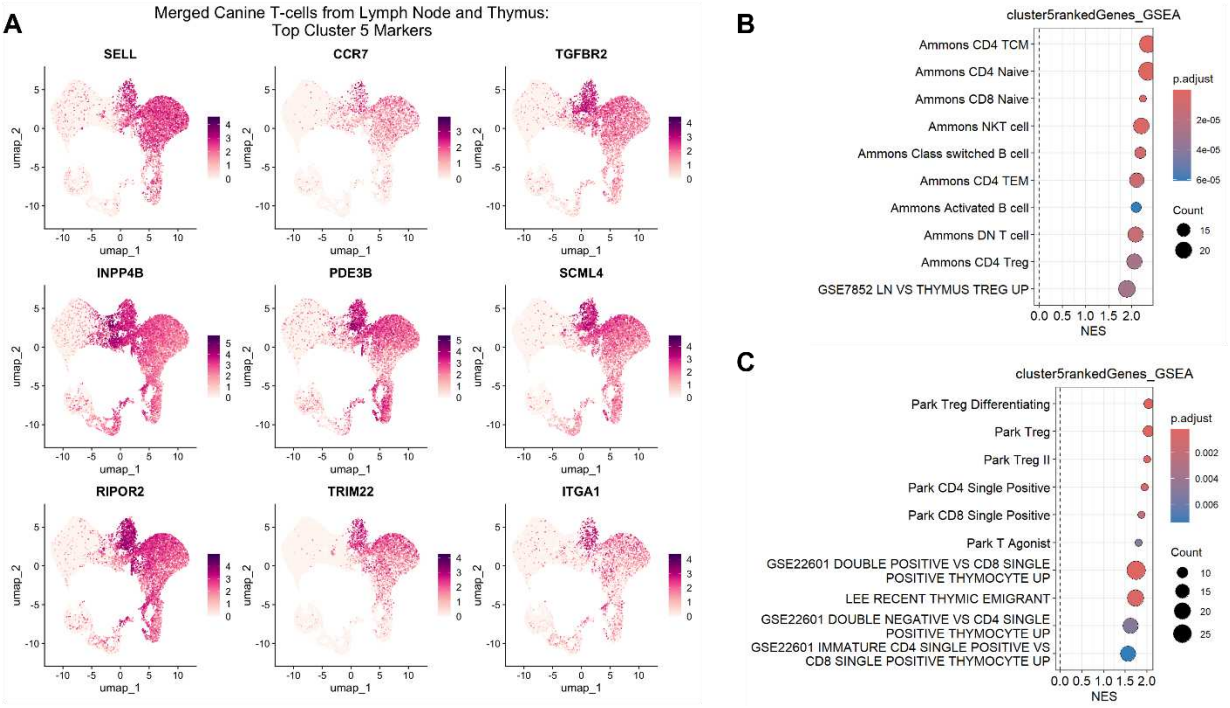


Figure 31: Annotation details of cluster 5. (**A**) Expression of key developmental markers and top differentially expressed markers in cluster 5. (**B-C**) GSEA enrichment results for gene sets derived from a human immune cell atlas (Conde et al.), a canine circulating leukocyte atlas (Ammons et al.), and gene sets of naïve and activated T-cell subset gene signatures from the Broad Institute Molecular Signatures Database (**B**) and derived from a human single-cell thymus atlas (Park et al.) and sets of genes upregulated by various T-cell progenitor and thymocyte subsets from the Broad Institute Molecular Signatures Database (**C**).

Cluster 6 of our merged atlas, originally derived from the thymus, was characterized by strong expression of cell cycle genes and enrichment for gene sets of proliferating DP and DN thymocytes and intrathymic T progenitors by GSEA (Table 7, Figure 32). Many of the top markers expressed by this cluster were also highly expressed by DP thymocytes in clusters 1 and 3. Thymocytes undergo a proliferative burst after β-selection between the late DN and early DP stages, and we suspect this highly proliferative phenotype is represented in cluster 6.

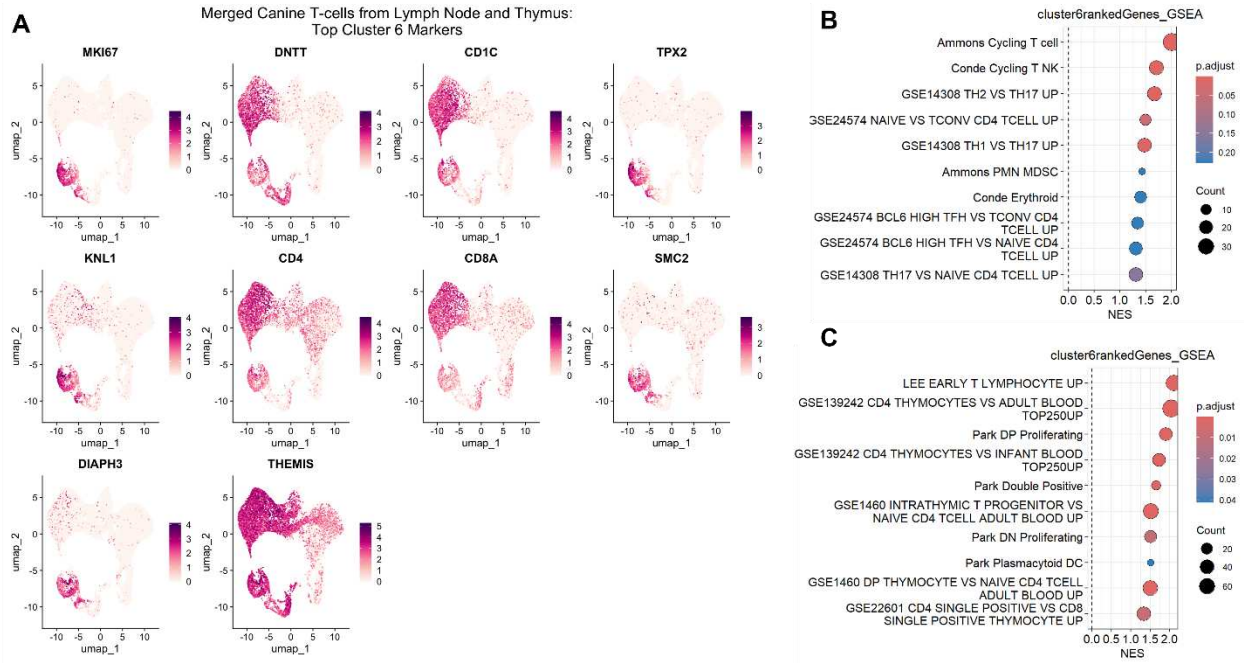


Figure 32: Annotation details of cluster 6. (**A**) Expression of key developmental markers and top differentially expressed markers in cluster 6. (**B-C**) GSEA enrichment results for gene sets derived from a human immune cell atlas (Conde et al.), a canine circulating leukocyte atlas (Ammons et al.), and gene sets of naïve and activated T-cell subset gene signatures from the Broad Institute Molecular Signatures Database (**B**) and derived from a human single-cell thymus atlas (Park et al.) and sets of genes upregulated by various T-cell progenitor and thymocyte subsets from the Broad Institute Molecular Signatures Database (**C**).

Cluster 7 of our merged atlas, originally derived from the lymph node, was characterized by expression of markers associated with $CD8^+$ T cells and NK cells, including *KLRB1, KLRK1, CD96,* and *GZMA*, and similarly was enriched for gene sets associated with these cell types (Table 7, Figure 33).

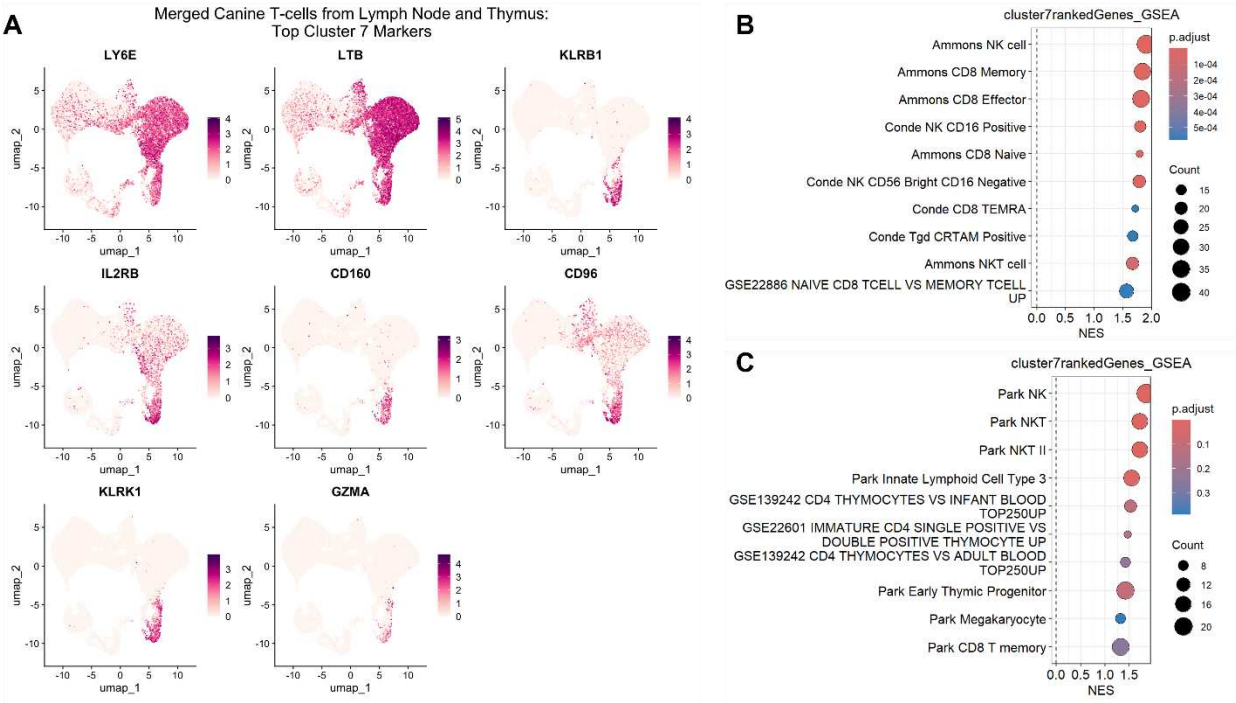


Figure 33: Annotation details of cluster 7. (**A**) Expression of key developmental markers and top differentially expressed markers in cluster 7. (**B-C**) GSEA enrichment results for gene sets derived from a human immune cell atlas (Conde et al.), a canine circulating leukocyte atlas (Ammons et al.), and gene sets of naïve and activated T-cell subset gene signatures from the Broad Institute Molecular Signatures Database (**B**) and derived from a human single-cell thymus atlas (Park et al.) and sets of genes upregulated by various T-cell progenitor and thymocyte subsets from the Broad Institute Molecular Signatures Database (**C**).

Cluster 8 of our merged atlas, originally derived from the thymus, was characterized by expression of several markers of stemness, including *CD34, KIT, DNTT,* and *ALPL*, that could indicate the presence of early lymphoid progenitor cells in this cluster (Figure 34A). Notably, this cluster contained a subset of cells that did not express either *CD4* or *CD8A*, which may represent DN thymocytes, and a subset of cells that express *CD8A* but not *CD4,* which may represent the $CD8^{+}CD4^{-}$ immature single positive (ISP) stage that occurs between the DN4 and DP stages of thymocyte development following β-selection (B). This cluster was significantly enriched for a gene set of genes upregulated in ISP thymocytes compared to DP thymocytes, further supporting the representation of this phenotype by cells in cluster 8 (Table 7, Figure 34D). Cluster 8 also exhibited the highest expression of *CDH1*, the gene encoding E-cadherin, of any cluster in our atlas (Figure 34A). In fetal and neonatal mice, E-cadherin is expressed by thymocytes—with the highest

expression seen on the most immature DN thymocytes—and is thought to play a role in mitigating the early interactions between thymocytes and thymic stromal cells.[53] E-cadherin expression on murine thymocytes typically diminishes significantly beyond the neonatal stage, and its expression has not been demonstrated on human thymocytes at any developmental stage.[54] The significance of *CDH1* expression in these samples from 7- to 9-month-old dogs is therefore unclear but may further support this cluster representing the earliest DN thymocytes.

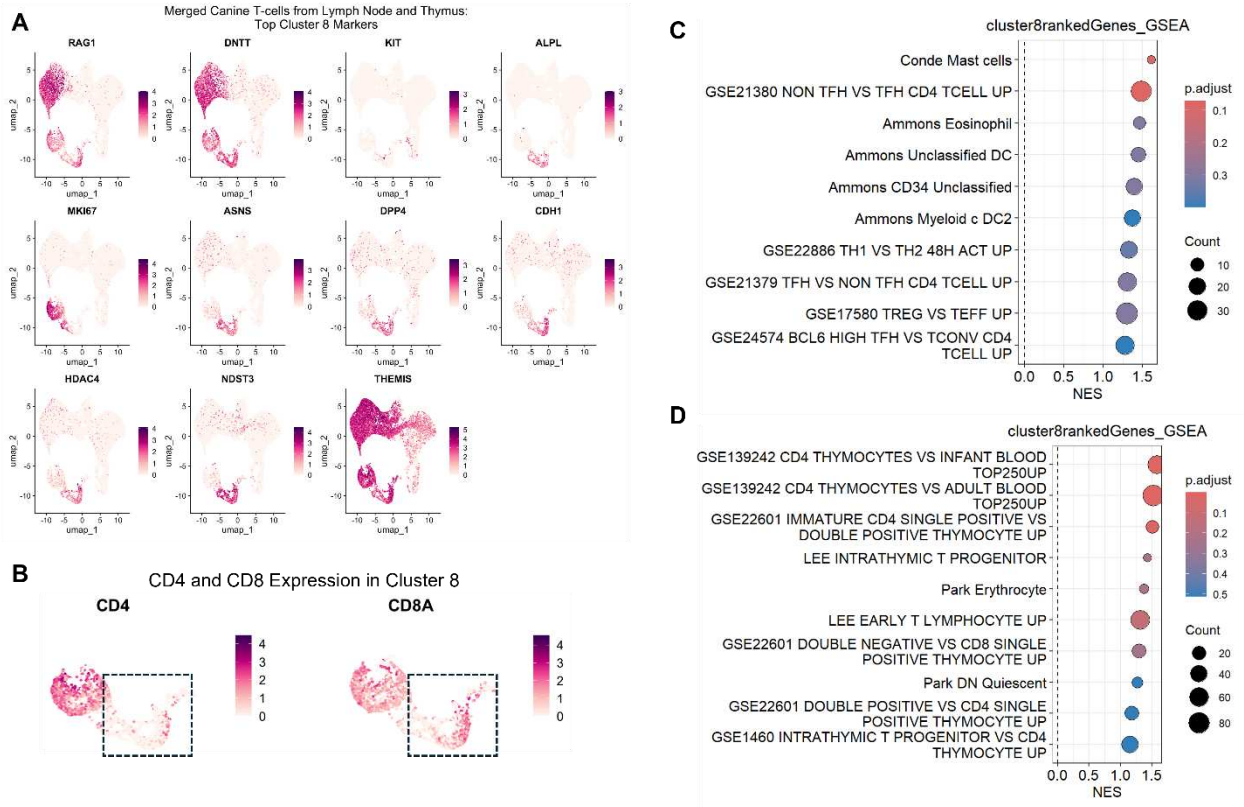


Figure 34: Annotation details of cluster 8. (**A**) Expression of key developmental markers and top differentially expressed markers in cluster 8. (**B**) Within cluster 8 is a subset of $CD4^{-}CD8^{-}$ (DN) cells and a subset of $CD4^{-}CD8^{+}$ cells that may represent the ISP stage of thymocyte development. (**C**) No gene sets derived from a human immune cell atlas (Conde et al.), a canine circulating leukocyte atlas (Ammons et al.), and gene sets of naïve and activated T-cell subset gene signatures from the Broad Institute Molecular Signatures Database were significantly enriched in cluster 8. (**D**) GSEA enrichment results for gene sets derived from a human single-cell thymus atlas (Park et al.) and sets of genes upregulated by various T-cell progenitor and thymocyte subsets from the Broad Institute Molecular Signatures Database. "GSE139242 CD4 THYMOCYTES VS NFANT BLOOD TOP250UP", "GSE139242 CD4 THYMOCYTES VS ADULT BLOOD TOP250UP", and "GSE22601 IMMATURE CD4 SINGLE POSITIVE VS DOUBLE POSITIVE THYMOCYTE UP" were significantly enriched (padj < 0.05).

Cluster 9 of our merged atlas, which was originally part of the late SP2 cluster of the thymus-only atlas, was characterized by expression of primarily $CD8^{+}$ T and NK cell-associated

genes like *KLRB1* and *CD7*, as well as other markers of T-cell maturity like *MAF* and T-cell deactivator *LGALS3* (Figure 35A). This cluster was not significantly enriched for any thymocyte or more mature T-cell gene sets (Figure 35B-C); however, the late SP2 cluster it was originally derived from in the thymus was enriched for NK, NKT, Treg, and memory T cell gene sets. These findings, and the close distance in UMAP space of this cluster to lymph node-derived cluster 7 that was enriched for similar cell types, could indicate that this cluster is comprised of late SP cells polarized toward the $CD8^+$ lineage.

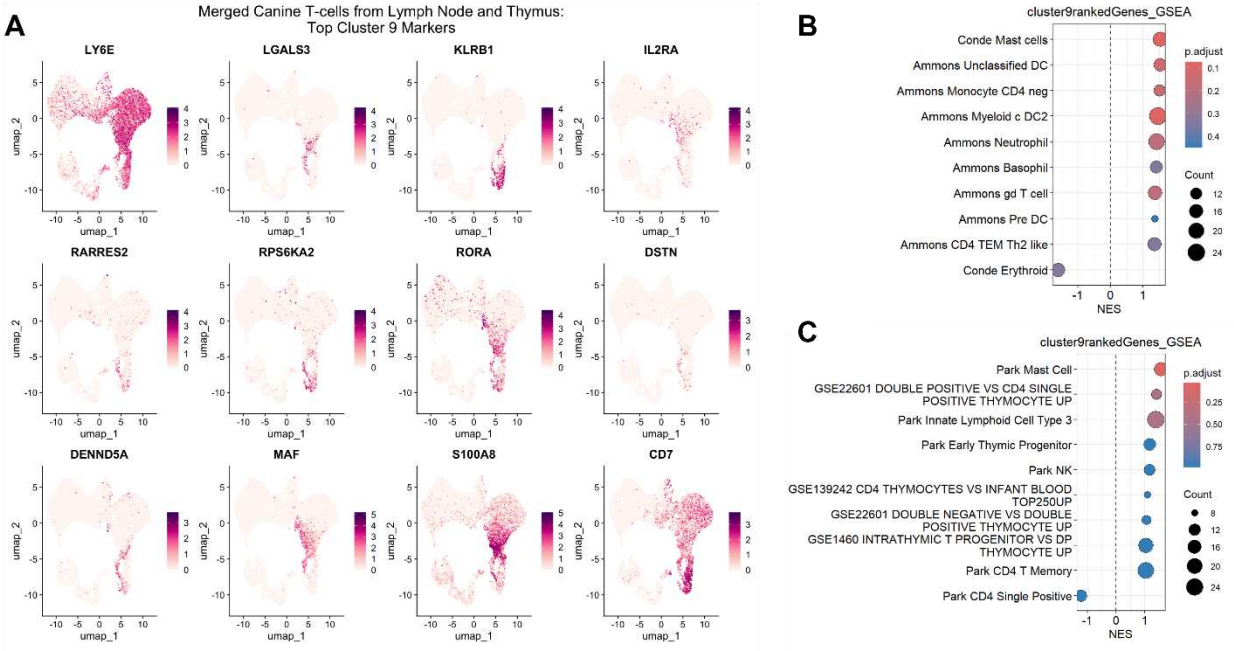


Figure 35: Annotation details of cluster 9. (**A**) Expression of key developmental markers and top differentially expressed markers in cluster 9. (**B**) No gene sets derived from a human immune cell atlas (Conde et al.), a canine circulating leukocyte atlas (Ammons et al.), and gene sets of naïve and activated T-cell subset gene signatures from the Broad Institute Molecular Signatures Database were significantly enriched in cluster 9. (**C**) GSEA enrichment results for gene sets derived from a human single-cell thymus atlas (Park et al.) and sets of genes upregulated by various T-cell progenitor and thymocyte subsets from the Broad Institute Molecular Signatures Database. Only "Park Mast Cell" was significantly enriched (padj < 0.05).

Cluster 10 of our merged atlas, which was originally part of the LateSP2 cluster of the thymus-only atlas, was contradictorily enriched for immature DN thymocyte, regulatory T cell, and γ/δ gene signatures when compared to existing human and canine single-cell atlases (Table 7, Figure 36B-C). The top differentially expressed genes characterizing this cluster included *HAVCR1*

and *HAVCR2*, which are typically expressed by more mature effector T cells;[55,56] *BCL2*, which is expressed in proliferating DN thymocytes, downregulated for the remaining stages of thymocyte development in the cortex, upregulated in SP thymocytes to promote their survival prior to thymic egress, and also expressed by CD8$^+$ memory T cells;[57–59] *RBPMS*, whose role in T-cell differentiation and development is poorly characterized; and *SOX5,* which drives Th17 cell differentiation (Figure 36A).[60,61] Given that the majority of these genes are associated with more mature T-cell subsets, we retained the previous annotation of late SP thymocyte for this cluster but acknowledge that this contradicts the enrichment of this cluster for human immature DN thymocyte gene signatures and may not appropriately represent the biologic cell type of this cluster.

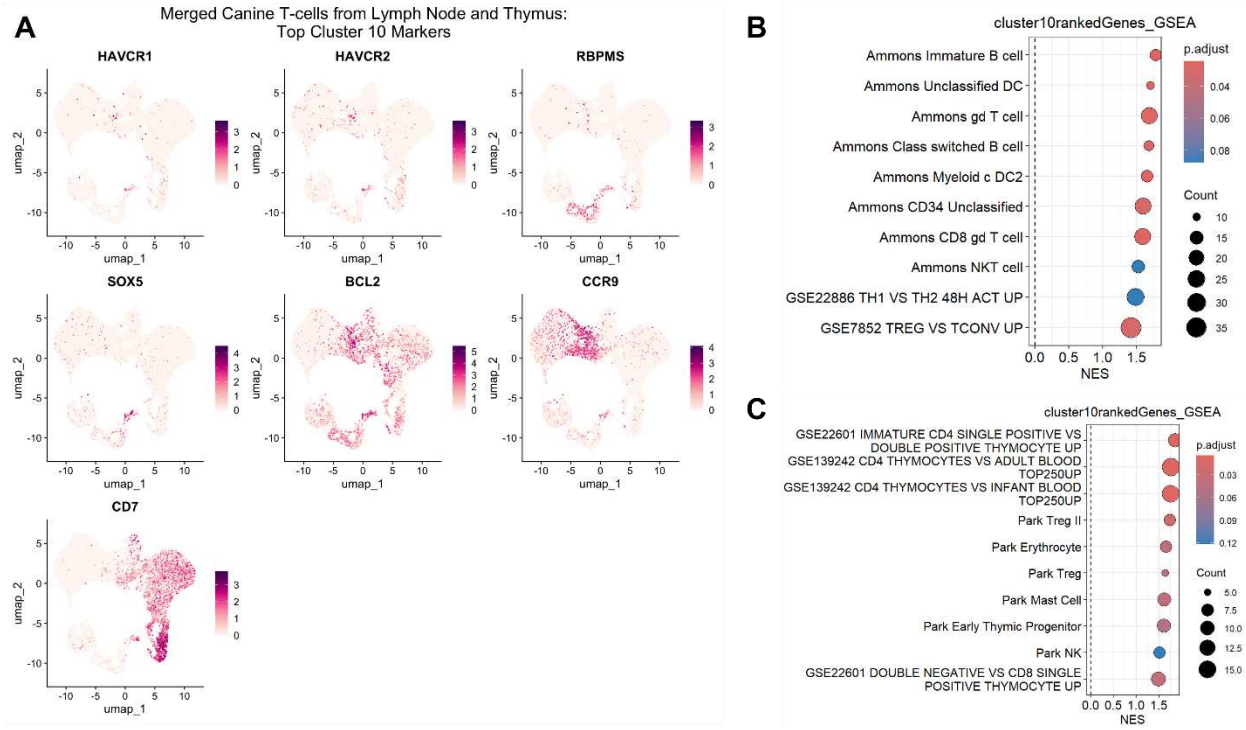


Figure 36: Annotation details of cluster 10. (**A**) Expression of key developmental markers and top differentially expressed markers in cluster 10. (**B-C**) GSEA enrichment results for gene sets derived from a human immune cell atlas (Conde et al.), a canine circulating leukocyte atlas (Ammons et al.), and gene sets of naïve and activated T-cell subset gene signatures from the Broad Institute Molecular Signatures Database (**B**) and derived from a human single-cell thymus atlas (Park et al.) and sets of genes upregulated by various T-cell progenitor and thymocyte subsets from the Broad Institute Molecular Signatures Database (**C**).

The final annotations assigned to the clusters in our merged T-cell lineage atlas prior to pseudotemporal trajectory analysis are indicated in Figure 37 and summarized in Table 7.

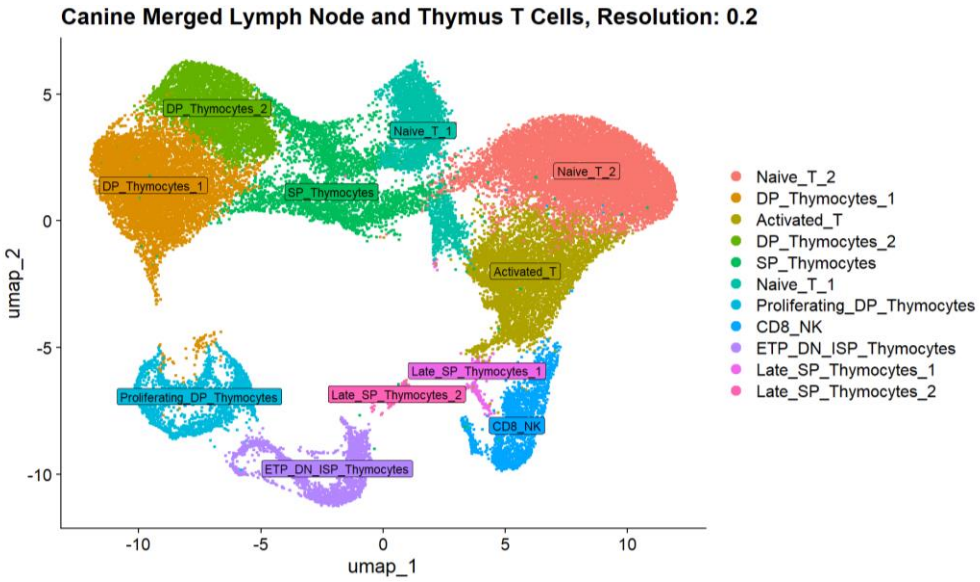


Figure 37: Final annotations assigned to single-cell clusters of normal canine thymus samples prior to subsetting T cell clusters and merging with canine lymph node single-cell data.

Table 7: Summary of cluster annotation for the merged canine thymus and lymph node single-cell RNA-seq data.

Cluster	Original source	Expressed markers	Summary of enriched gene sets	Final assignment
0	Lymph node 0	*CD3, CD4, CD8A, SELL, CCR7, LY6E, LTB, HLA-DQB2, TRIM22, DLA-DRA*	Naïve $CD4^+$ and $CD8^+$ T cells	Naïve T cells 2 (Naïve_T_2)
1	Thymus 0	*CD3, RAG1, DNTT, CD4, CD8A, CD1C, PDE1C, NT5DC2, THEMIS*	DP and quiescent DP thymocytes	Double positive thymocytes 1 (DP_Thymocytes_1)
2	Lymph node 2	*CD3, CD4, CD8A, SELL, LY6E, CTLA4, LTB, IL2RA, MAF, S100A8, ICOS*	Th1, Th2, Th17, Treg, Trm, and Tem cells	Activated T cells (Activated_T)
3	Thymus 1	*CD3, RAG1, DNTT, CD4, CD8A, CD1C, SLC7A11, DBX2, ADAM12, ARRP21, PDE1C, THEMIS*	DP thymocytes and quiescent DP thymocytes, early T lymphocytes, intrathymic progenitor cells	Double positive thymocytes 2 (DP_Thymocytes_2)
4	Thymus 2	*CD3, CD4, CD8A, CCR9, ASAP1, CELF2, TOX2, GATA3, THEMIS*	CD4 SP thymocytes	Single positive thymocytes (SP_Thymocytes)

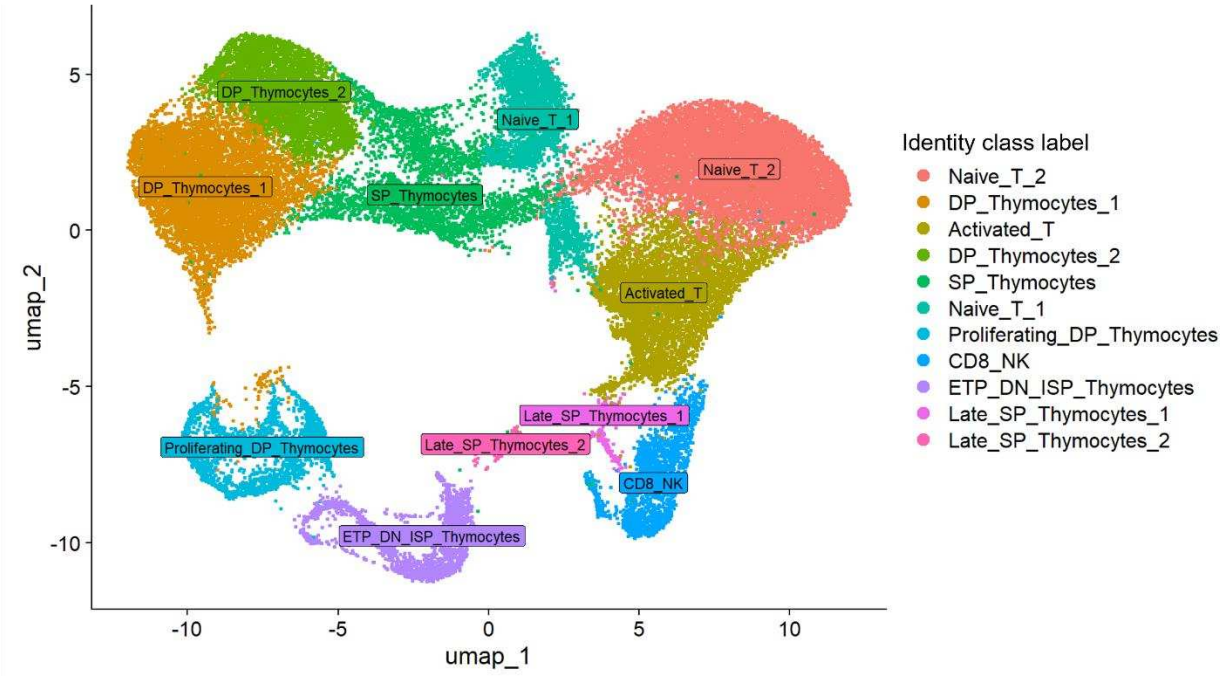
5	Lymph node 3	*CD3, CD4, CD8A, SELL, CCR7, TGFBR2, INPP4B, PDE3B, SCML4, RIPOR2, TRIM22, ITGA1*	Naïve $CD4^+$ and $CD8^+$ T cells, Tregs, γδ Trm cells, Tcm cells, Tem cells	Naïve T cells 1 (Naïve_T_1)
6	Thymus 4 and Thymus 8	*CD3, DNTT, CD4, CD8A, MKI67, CD1C, TPX2, KNL1, SMC2, DIAPH3, THEMIS*	Proliferating DP and DN thymocytes, intrathymic T progenitors	Proliferating double positive thymocytes (Proliferating_DP_Thymocytes)
7	Lymph node 4	*CD3, CD4, CD8A, LY6E, LTB, KLRB1, IL2RB, CD160, CD96, KLRK1, GZMA*	NK cells, NKT cells, $CD8^+$ T memory and effector cells, $CD8^+$ effector cells, naïve $CD8^+$ T cells	CD8 T / NK cells (CD8_NK)
8	Thymus 3	*CD3, CD34, KIT, ALPL, RAG1, DNTT, MKI67, ASNS, DPP4, CDH1, HDAC4, NDST3, THEMIS*	Immature CD4 SP thymocytes (DN to DP transition stage)	Early thymic progenitors, double negative thymocytes, and immature single positive thymocytes (ETP_DN_ISP_Thymocytes)
9	Thymus 6	*CD3, CD4, CD8A, LY6E, LGALS3, KLRB1, IL2RA, RARRES2, RPS6KA2, RORA, DSTN, DENND5A, MAF, S100A8, CD7*	NK cells, NKT cells, ILC3 cells, Tregs, $CD4^+$ and $CD8^+$ memory T cells, early thymic progenitors	Late single positive thymocytes 1 (Late_SP_Thymocytes_1)
10	Thymus 6	*CD3, CD4, CD8A, HAVCR1, HAVCR2, RBPMS, SOX5, BCL2, CCR9, CD7*	Tregs, γδ T cells, DN thymocytes, immature CD4 SP thymocytes (DN to DP transition stage)	Late single positive thymocytes 2 (Late_SP_Thymocytes_2)

Prior to trajectory analysis, clustering of our merged thymus and lymph node single-cell data was repeated with Monocle 3's Leiden community detection algorithm, and this clustering largely recapitulated the clusters identified in Seurat with a few exceptions. First, lymph node-derived cluster 5 (presumed naïve T cells) and thymus-derived cluster 4 (presumed SP cells) in our Seurat atlas were partially combined into a single cluster by Monocle 3 (cluster 5). Lymph node-derived clusters 0 (presumed naïve T cells) and 2 (presumed activated T cells) of our Seurat atlas were divided into 4 clusters by Monocle 3 (clusters 3, 4, 6, and 7), and thymus-derived cluster 9

(presumed late SP cells) of our Seurat atlas was partially absorbed into this new cluster 7 and partially absorbed into cluster 10 of the Monocle 3 atlas. New labels were assigned to these Monocle 3 clusters prior to downstream analysis that best preserved the previously assigned annotations (Figure 38).

A

Annotated Seurat clusters



B

Annotated Monocle 3 clusters

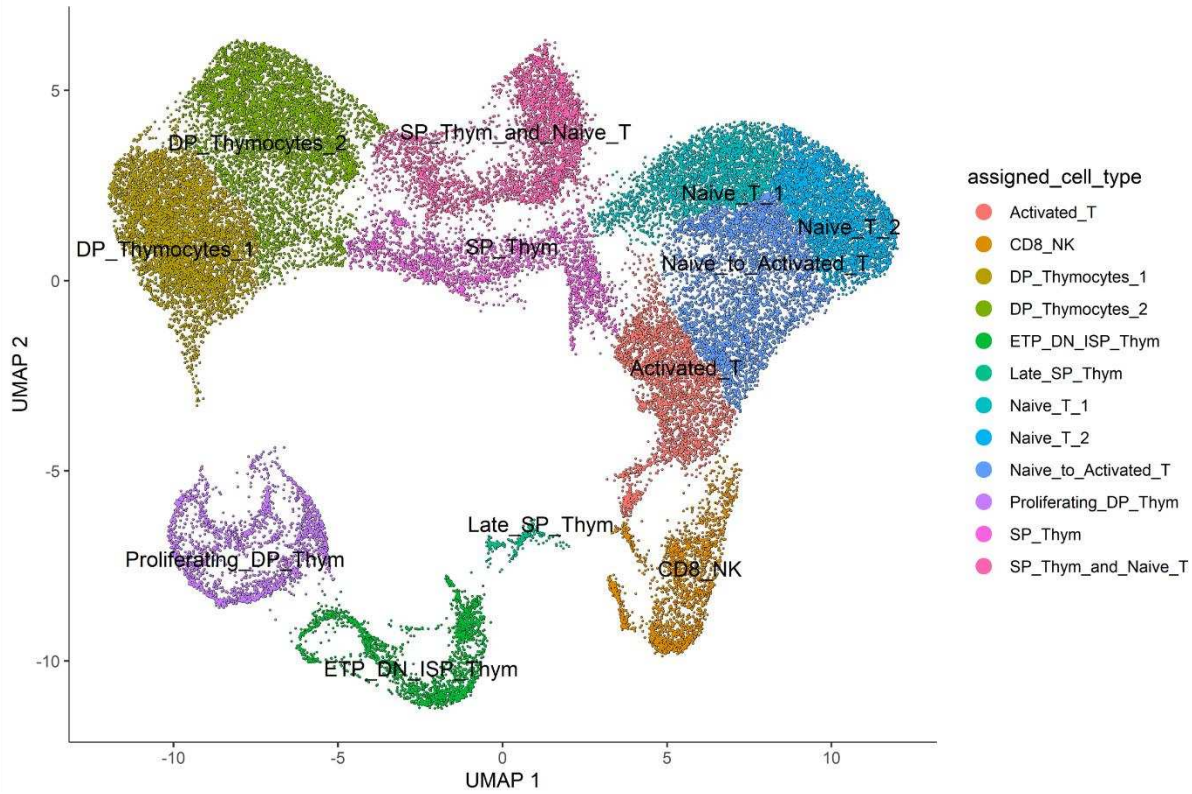
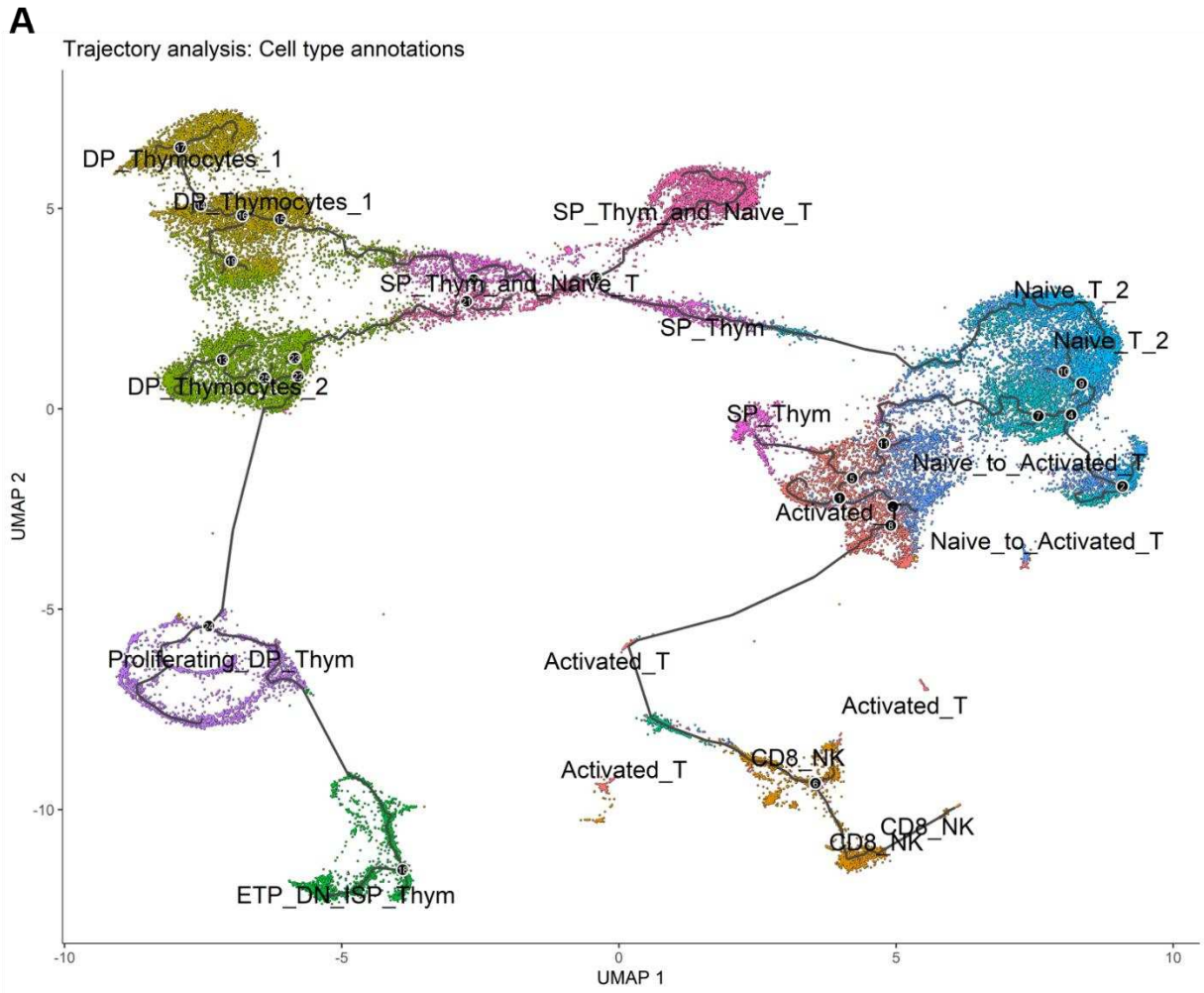
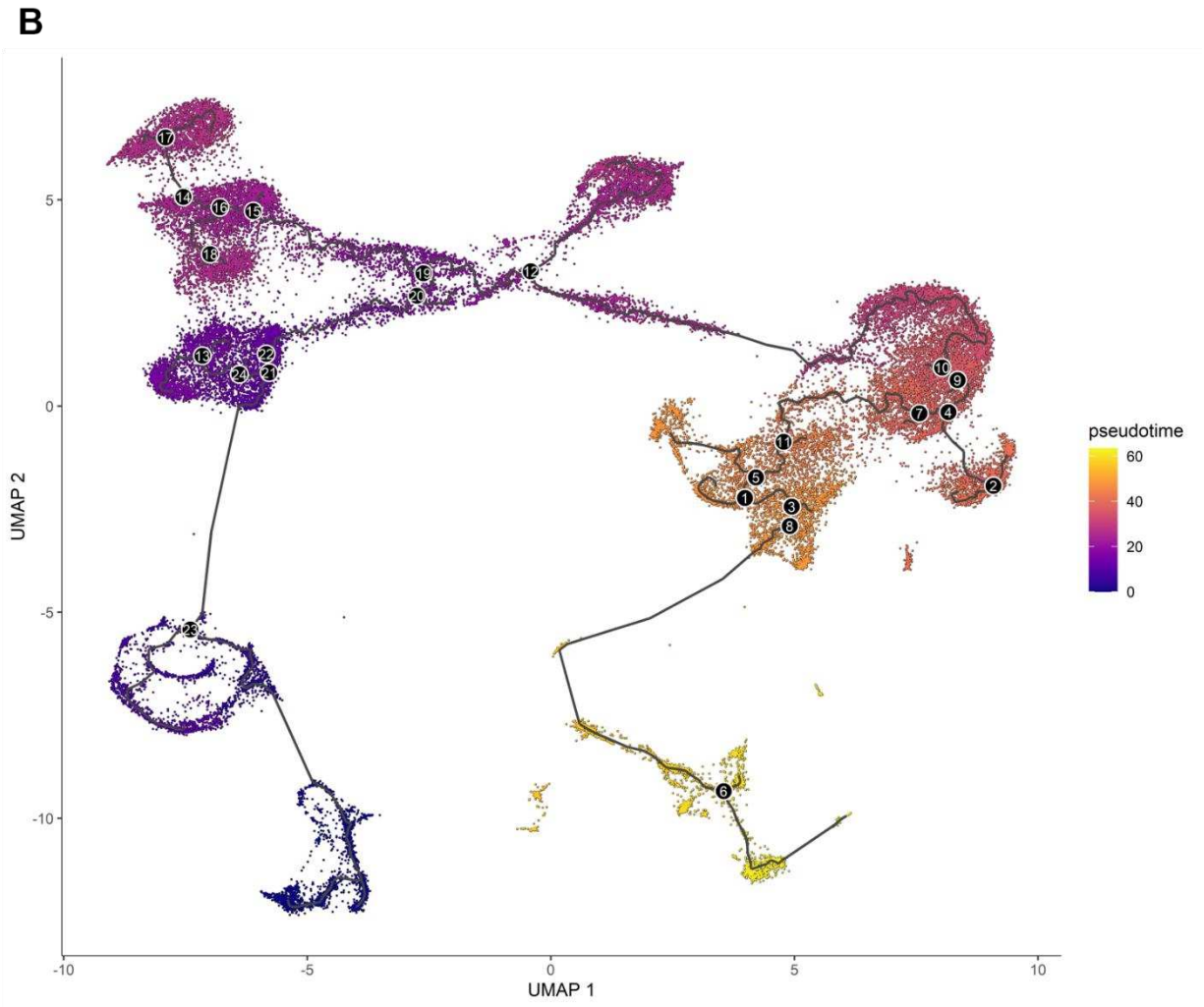


Figure 38: Comparison of Seurat::FindClusters() and monocle3::cluster_cells. Lymph node-derived cluster "Naïve_T_1" and thymus-derived cluster "SP_Thymocytes" of the Seurat atlas (**A**) were partially combined into a single cluster by Monocle 3 ("SP_Thym_and_Naive_T") (**B**). Lymph node-derived cluster "Naïve_T_2" and "Activated_T" of the Seurat atlas were divided into 4 distinct clusters by Monocle 3, subsequently annotated as "Naïve_T_1", "Naïve_T_2", "Naïve_to_Acitvated_T", and "Activated_T". Thymus-derived cluster "Late_SP_Thymocytes_1" of the Seurat atlas was partially absorbed into the "Activated_T" cluster and partially absorbed into the "CD8_NK" cluster of the Monocle 3 atlas.

Pseudotime trajectory analysis performed with Monocle 3 successfully recapitulated the expected timeline of T-cell differentiation based on our manual cluster annotations, with the cluster of presumed early thymic progenitors, DN thymocytes, and ISP thymocytes assigned to the earliest point in pseudotime and clusters of presumed activated nodal T-cell subsets assigned to the latest points in pseudotime (Figure 39). Several branches representing key cell fate decision points were predicted, including branches at the DP to SP thymocyte transition, the SP to naïve T cell transition, and the naïve to activated nodal T cell transition (Figure 39B). Interestingly, the cluster encompassing both presumed SP cells derived from the thymus and naïve T cells derived from the lymph node ("SP_Thym_and_Naive_T") was placed similarly in pseudotime to one of the thymus-derived clusters of DP thymocytes ("DP_Thymocytes_1") (Figure 39C).


A
Trajectory analysis: Cell type annotations
DP_Thymocytes_1
DP_Thymocytes_1
SP_Thym_and_Naive_T
SP_Thym_and_Naive_T
SP_Thym
Naive_T_2
Naive_T_2
DP_Thymocytes_2
SP_Thym
Naive_to_Activated_T
Activated_T
Naive_to_Activated_T
Proliferating_DP_Thym
Activated_T
Activated_T
Activated_T
CD8_NK
CD8_NK
CD8_NK
ETP_DN_ISP_Thym
UMAP 2
UMAP 1
5
0
-5
-10
-10
-5
0
5
10


B
UMAP 2
UMAP 1
pseudotime
60
40
20
0

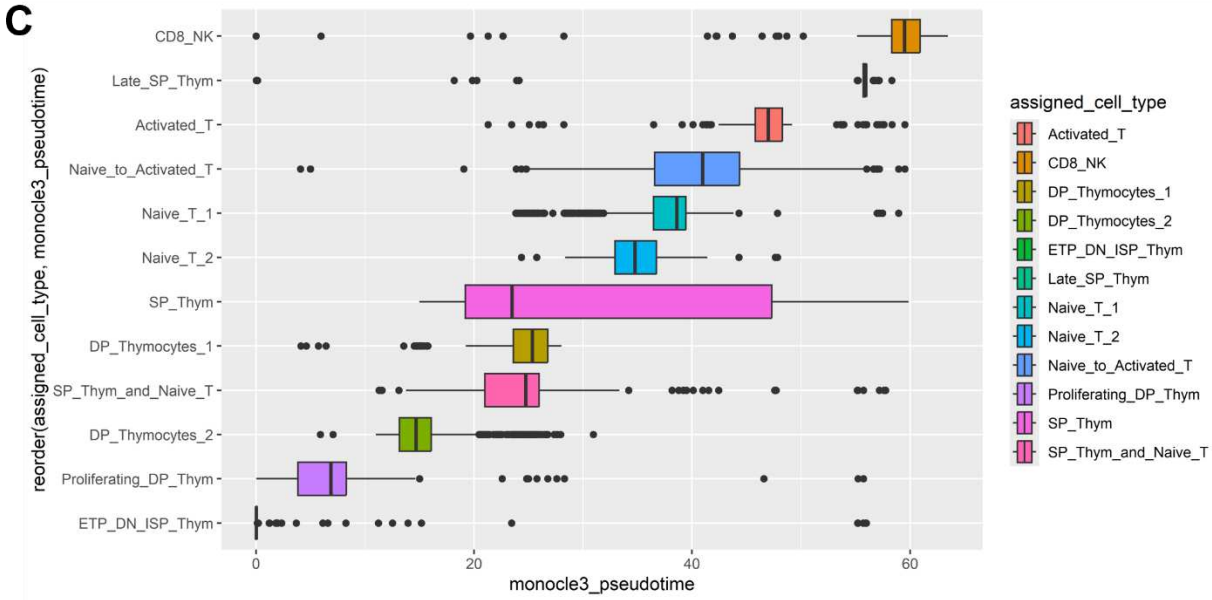

C
reorder(assigned_cell_type, monocle3_pseudotime)
CD8_NK
Late_SP_Thym
Activated_T
Naive_to_Activated_T
Naive_T_1
Naive_T_2
SP_Thym
DP_Thymocytes_1
SP_Thym_and_Naive_T
DP_Thymocytes_2
Proliferating_DP_Thym
ETP_DN_ISP_Thym
monocle3_pseudotime
assigned_cell_type
Activated_T
CD8_NK
DP_Thymocytes_1
DP_Thymocytes_2
ETP_DN_ISP_Thym
Late_SP_Thym
Naive_T_1
Naive_T_2
Naive_to_Activated_T
Proliferating_DP_Thym
SP_Thym
SP_Thym_and_Naive_T

Figure 39: Pseudotime trajectory analysis of the canine T-cell lineage across thymus and peripheral lymph node. (**A**) Learning the trajectory graph with monocle3::learn_graph(). Lines connecting clusters of our merged lymph node and thymus single-cell atlases represent edges of the minimum spanning tree constructed by Monocle 3. The longest path through the minimum spanning tree represents the longest sequence of transcriptionally similar cells and serves as the backbone for ordering cells in pseudotime. Paths that branch from this longest sequence are alternative trajectories representing cell fate decisions. In our data, major branch points included the DP to SP thymocyte transition, the SP to naïve T cell transition, and naïve to activated nodal T cell transition. (**B**) Cells ordered in pseudotime with monocle3::order_cells(). Dark purple represents the earliest points in pseudotime while yellow represents the latest. Numbered black circles indicate branch nodes, representing different cell fate outcomes along the trajectory. The numbers within the circles are for reference purposes only, with no significance to their order. (**C**) Box plots illustrating the distribution of cells belonging to each cluster of our single-cell atlas across pseudotime. The ordering of clusters in our canine single-cell data largely recapitulates the expected biologic timeline for differentiation of thymocytes and nodal T cells.

Differential expression analysis using Moran's *I* spatial autocorrelation and pseudo R-squared specificity scores identified several key genes whose expression changed across pseudotime (Figure 40, Supplemental Table 15), many of which helped validate cluster annotations. For example, studies in mice have shown that the gene *GIMAP4* is upregulated in DN4 thymocytes in response to pre-TCR signaling, then transiently downregulated in the DP thymocyte stages, and finally re-expressed in SP cells and peripheral T cells.[62] In our data, *GIMAP4* was identified as one of the most significantly changing genes across pseudotime, expressed in the ETP_DN_ISP_Thym cluster, to a lesser degree in the Proliferating_DP_Thym cluster, then not expressed in DP_Thymocytes_1 or DP_Thymocytes_2 before being expressed in our SP thymocyte clusters and nodal T cell clusters (Figure 40).

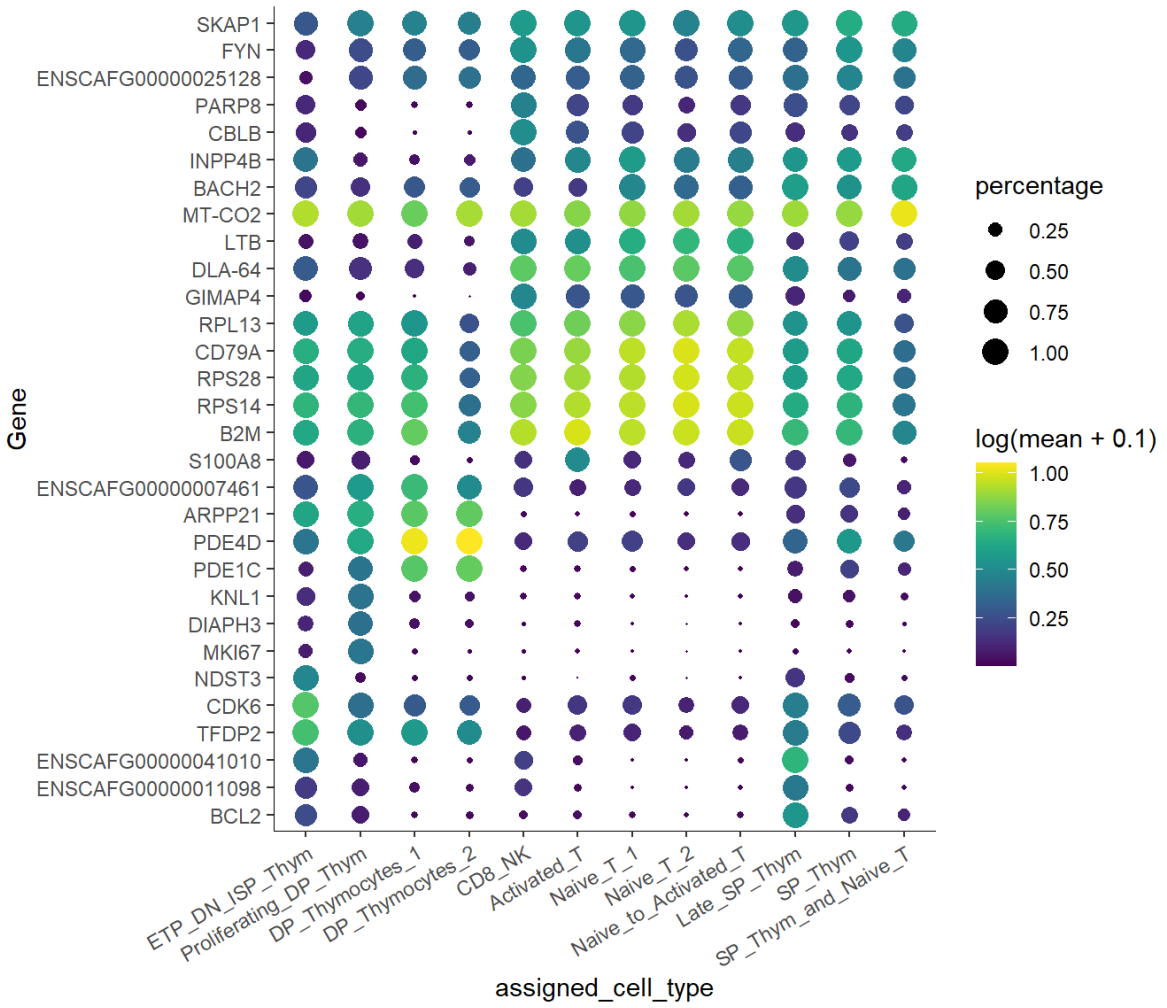
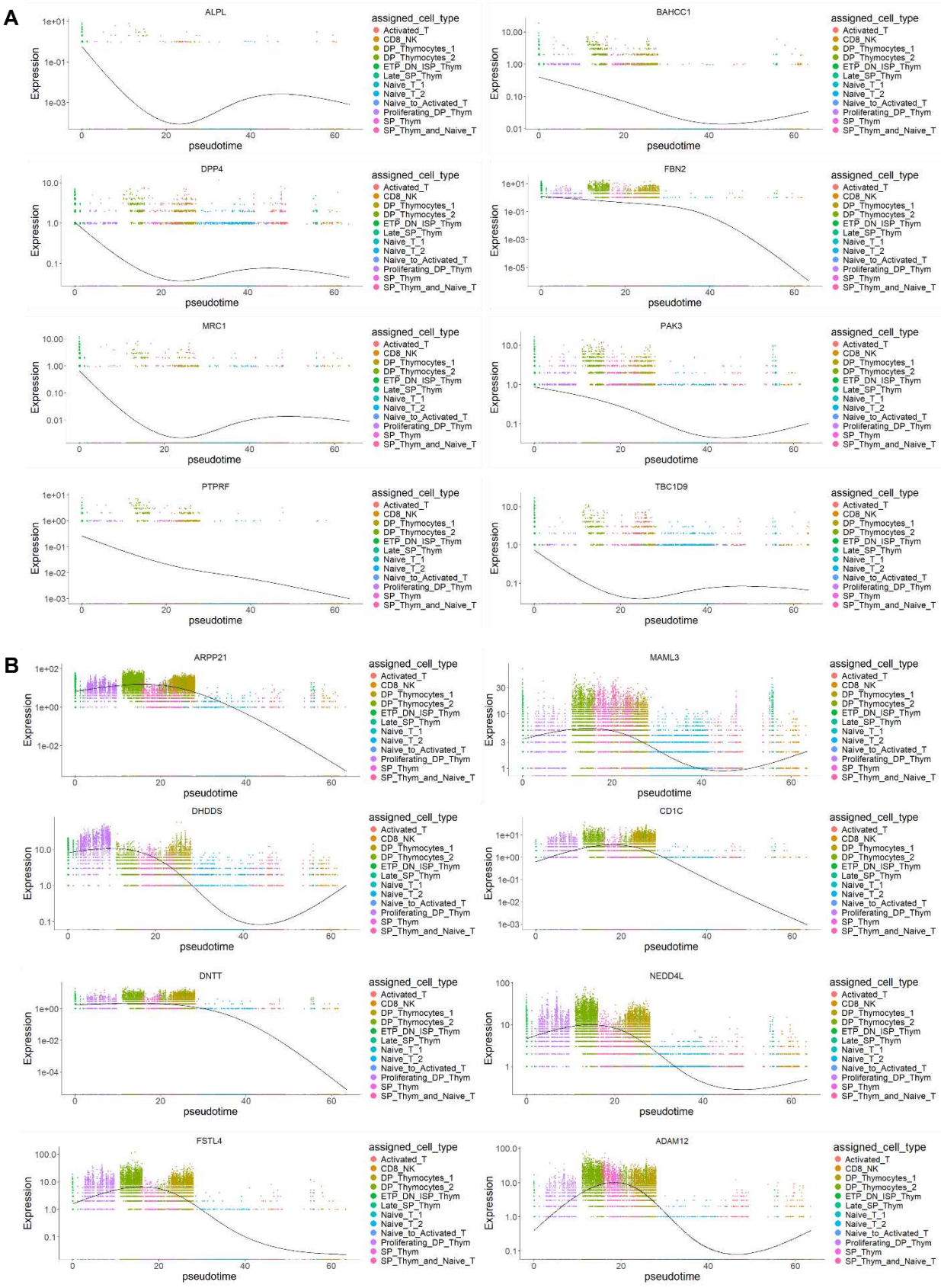


Figure 40: Top markers of each Monocle 3 cluster. The top 3 genes expressed by at least 85% of the cells in each cluster resulting from Monocle3::find_clusters(). Genes were ranked according to *pseudo R-squared* values, a specificity metric ranging from 0 to 1.

Other genes expressed at the earliest points in pseudotime included known markers of stemness *ALPL* and *TBC1D9*, which were most highly expressed in our ETP_DN_ISP_Thym cluster (Figure 41A), and canonical markers of early T-cell development, such as *DNTT*, which were highly expressed from early thymic progenitors to DP thymocytes in our atlas (Figure 41B). Similarly, genes turned on late in pseudotime included known markers of T-cell maturity, such as *DLA-DRA, SELL,* and *LTB* (Figure 41D). However, the key changing genes identified by our trajectory analysis also included several genes whose role in the T-cell lineage is not currently well known, such as the expression of *DHDDS* early in pseudotime and the many ribosomal proteins expressed later in

pseudotime (Figure 41D). Genes turned on around in the mid-range of T-cell developmental pseudotime included *RIPOR2* and *TGFBR2*, which as discussed previously are important for maintaining T-cell quiescence prior to activation; *BCL2,* which is not expressed in most cortical stages of thymocyte development but is rapidly upregulated in more mature medullary SP thymocytes to promote their survival prior to thymic egress;[59] *INPP4B,* a promoter of PI3K/AKT signaling that plays a role in maintaining innate lymphocyte and NK cell populations in peripheral tissues, including the lymph nodes;[63] and *TOX2* and *SCML4*, whose roles in T-cell development were discussed previously (Figure 41C).


A
ALPL
BAHCC1
DPP4
FBN2
MRC1
PAK3
PTPRF
TBC1D9
B
ARPP21
MAML3
DHDDS
CD1C
DNTT
NEDD4L
FSTL4
ADAM12
Expression
pseudotime
assigned_cell_type
Activated_T
CD8_NK
DP_Thymocytes_1
DP_Thymocytes_2
ETP_DN_ISP_Thym
Late_SP_Thym
Naive_T_1
Naive_T_2
Naive_to_Activated_T
Proliferating_DP_Thym
SP_Thym
SP_Thym_and_Naive_T

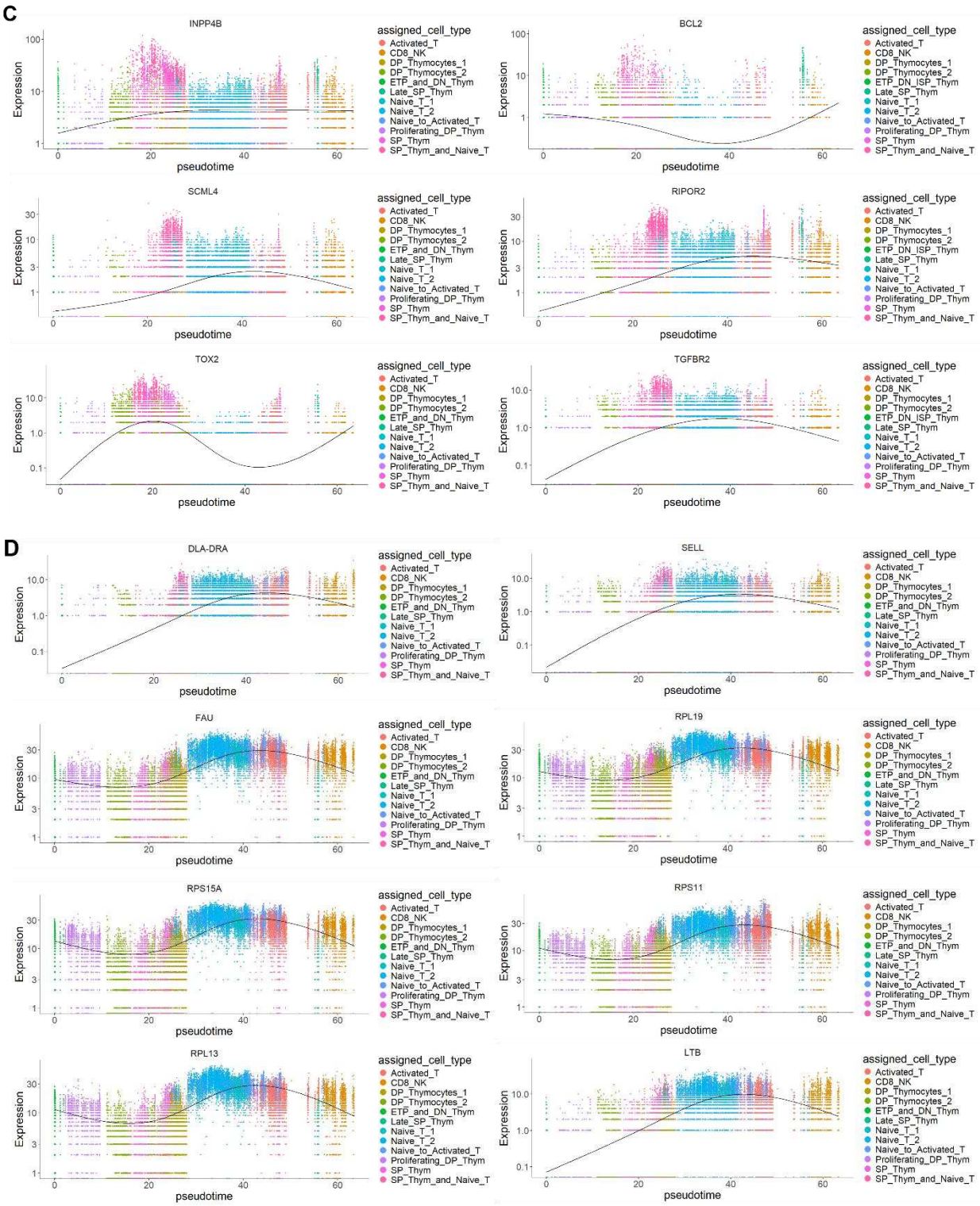


Figure 41: Changes in gene expression across pseudotime. (**A**) Snapshot of genes "on" earliest in pseudotime. Includes known markers of stemness *ALPL, BAHCC1,* and *TBC1D9*, as well as *DPP4, FBN2, MRC1, PAK3,* and *PTPRF*. (**B**) Snapshot of genes "on" around the DN/DP stage in pseudotime. Includes known early thymocyte marker *DNTT,* activator of Notch signaling *MAML3*, and *DHDDS, CD1C, NEDD4L, FSTL4,* and *NPR3*. (**C**) Snapshot of genes turned "on" in SP thymocytes and naïve T cells. Includes *RIPOR2* and *TGFBR2*, which maintain quiescence of T cells prior to activation; *TOX2*, known to be upregulated in thymocytes following positive selection; *BCL2,* which is upregulated to promote mature

SP thymocyte survival prior to thymic egress; as well as *INPP4B* and *SCML4*. (**D**) Snapshot of genes "on" late in pseudotime. Includes known markers of T-cell maturity *DLA-DRA, SELL,* and *LTB*, as well as *FAU, RPL19, RPS15A, RPS11,* and *RPL13*.

A subset of canine CD4$^+$ PTCL is enriched for gene signatures of immature canine thymocytes

Module scoring and GSVA were performed to determine which cell type(s) in our reference single-cell atlas of normal canine thymocytes and nodal T cells most resembled CD4$^+$ PTCL at a global gene expression level. The Seurat::AddModuleScore() function calculates the relative expression level of each cluster in a single-cell atlas for a list of features. To see which clusters had the most similar gene expression to canine CD4$^+$ PTCL, to our control nodal CD4$^+$ T cells, and to our control CD4$^+$ thymocytes, the list of all significantly upregulated genes in canine CD4$^+$ PTCL compared to control nodal CD4$^+$ lymphocytes, the list of all significantly upregulated genes in control nodal CD4$^+$ lymphocytes compared to CD4$^+$ PTCL, and the list of all significantly upregulated genes in sorted CD4$^+$ thymocytes compared to sorted CD4$^+$ nodal lymphocytes, respectively, were provided as feature sets for module scoring of our transcriptomic atlas clusters. Significantly upregulated genes were defined as those with a log2fc > 1 and padj < 0.05, according to differential expression analysis with DESeq2. Cell cycle genes were removed prior to these analyses, as cancer cells and progenitor cells are both known to be highly proliferative, and we wanted to investigate other shared genes that could more specifically inform cell of origin outside of this shared proliferative phenotype.

By module scoring, most of the cells that had higher relative expression of the list of genes significantly upregulated in canine CD4$^+$ PTCL compared to control nodal CD4$^+$ lymphocytes were dispersed throughout the DP_Thymocytes_2 and SP_Thymocytes clusters of our transcriptomic atlas of normal canine thymus and lymph node (Figure 42B). In contrast, most of the cells that had higher relative expression of genes significantly upregulated in control nodal CD4$^+$ lymphocytes compared to CD4$^+$ PTCL belonged to the Naïve_T_1 cluster of our atlas, with fewer scattered

throughout the Naïve_T_2 and Activated_T clusters (Figure 42C). The list of genes significantly upregulated in control CD4⁺ thymocytes compared to control nodal CD4⁺ lymphocytes from our bulk RNA-seq study had the highest relative enrichment in most of our thymus-derived single-cell clusters (Figure 42D).

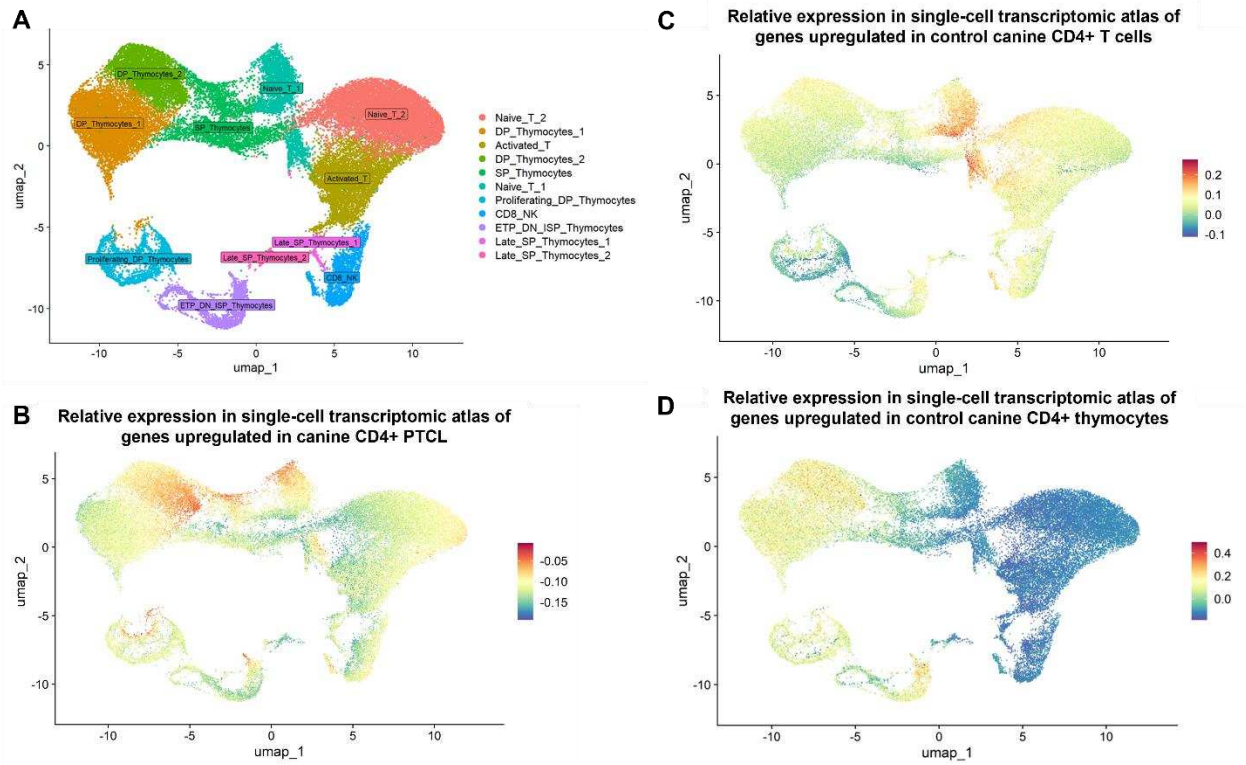


Figure 42: Module scoring using CD4⁺ PTCL, control CD4⁺ nodal lymphocyte, and control CD4⁺ thymocyte gene signatures. (**A**) Cluster annotations for our single-cell transcriptomic atlas of merged thymocyte and T-cell populations from normal canine thymus and lymph node samples. (**B**) Genes significantly upregulated (log2fc > 1 and padj < 0.05) in canine CD4⁺ PTCL compared to control nodal CD4⁺ lymphocytes had the highest relative expression in the DP_Thymocytes_2 and SP_Thymocytes clusters of our atlas, with fewer cells in the Naïve_T_1 and ETP_DN_ISP_Thymocytes clusters. (**C**) Genes significantly upregulated (log2fc > 1 and padj < 0.05) in control nodal CD4⁺ lymphocytes compared to CD4⁺ PTCL samples had the highest relative expression in the Naïve_T_1 cluster of our transcriptomic atlas, and in scattered cells of the Naïve_T_2 and Activated_T clusters. (**D**) Genes significantly upregulated (log2fc > 1 and padj < 0.05) in control CD4⁺ thymocytes compared to control nodal CD4⁺ lymphocytes had the highest relative expression in the thymus-derived clusters of our transcriptomic atlas.

By GSVA, CD4⁺ PTCLs of Cohort 1 exhibited stronger enrichment for the gene signatures of early thymocyte clusters of our transcriptomic atlas compared to gene signatures of naïve and activated T cell clusters of our atlas compared to control nodal CD4⁺ lymphocytes (Figure 43A). In Cohort 2, there appeared to be a subset of CD4⁺ PTCL that demonstrated some enrichment for naïve and activated T-cell gene signatures from our atlas, although this subset was frequently still

enriched for gene signatures of our more immature thymocyte gene signatures, as well (Figure 43B). A second cluster of CD4⁺ PTCL in Cohort 2 displayed only strong enrichment for the gene signature of proliferating DP thymocytes. The control nodal CD4⁺ thymocytes of Cohort 2 showed the strongest enrichment for DP thymocyte and naïve T cell signatures from our atlas.

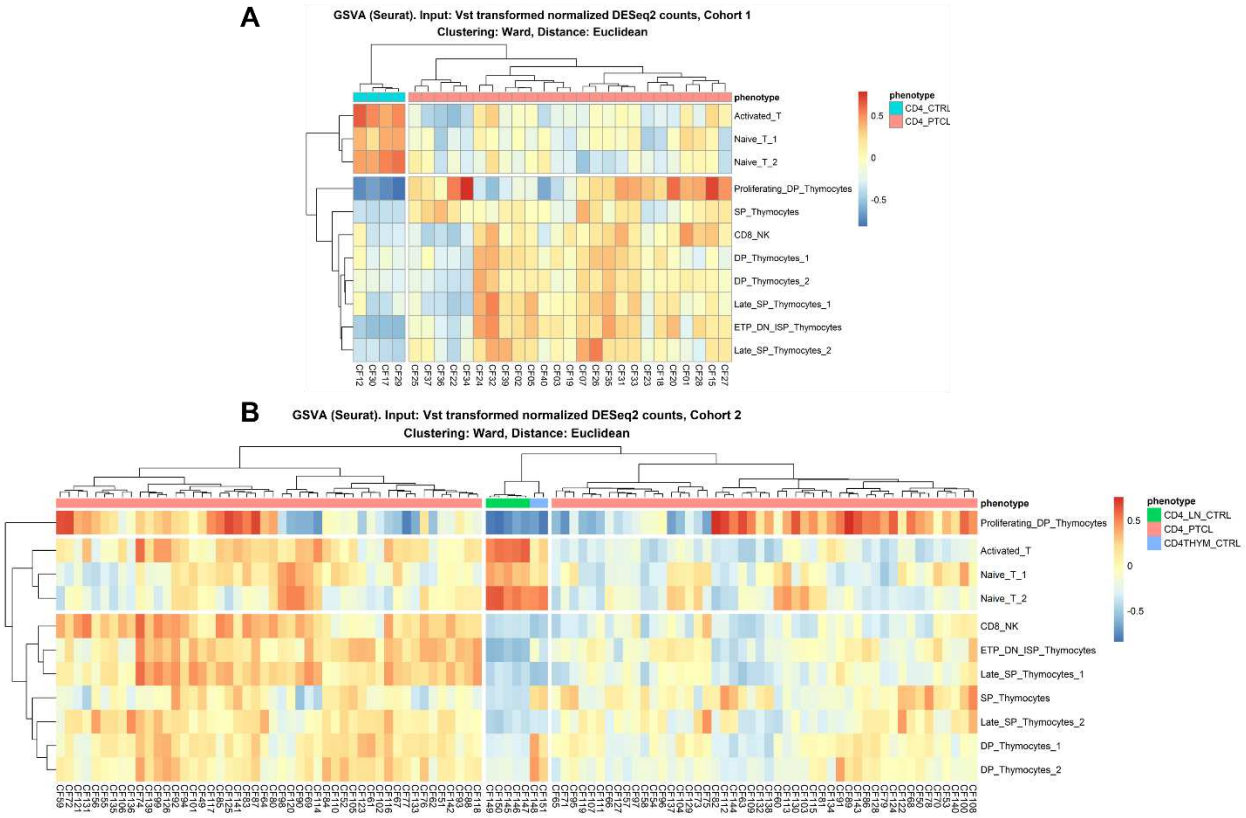
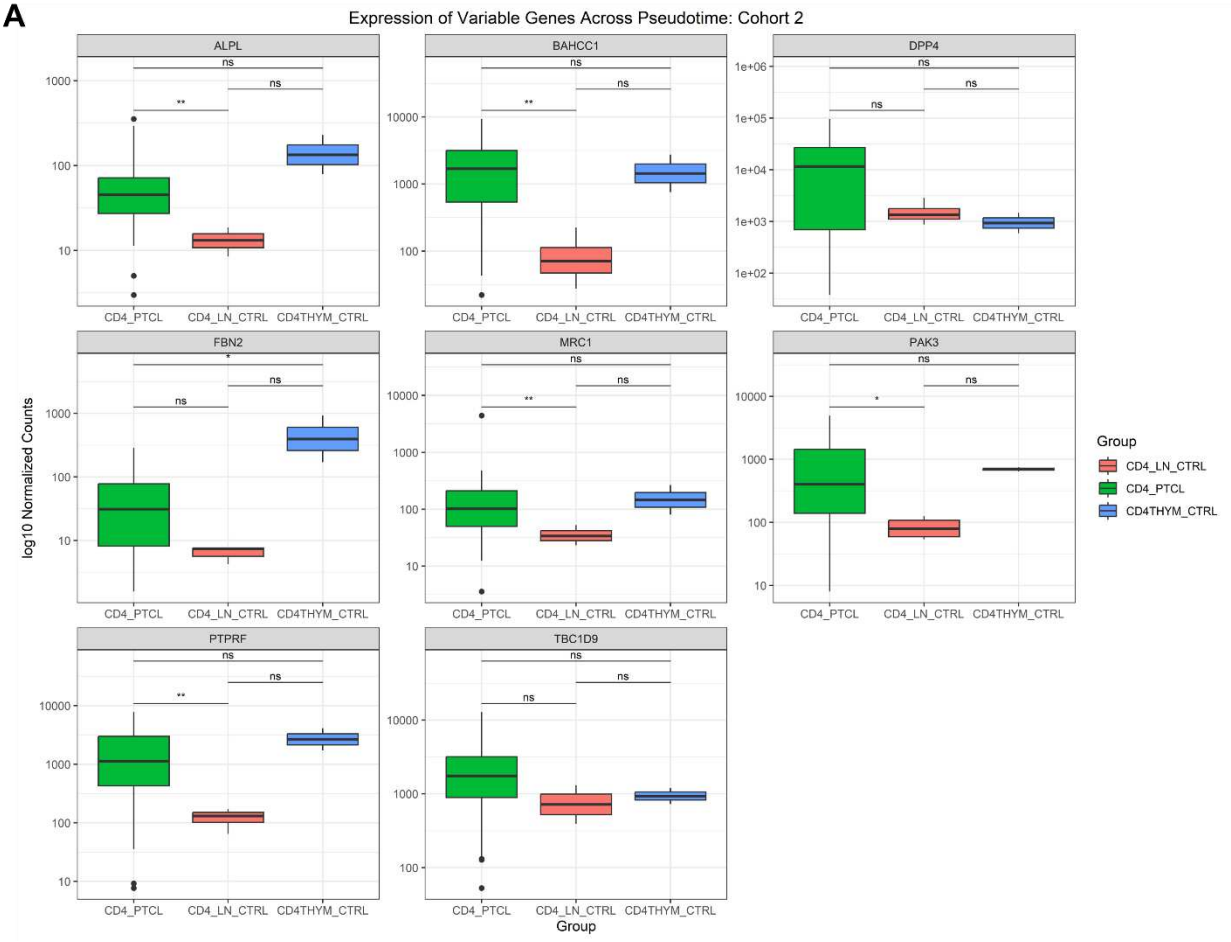


Figure 43: GSVA comparing canine CD4⁺ PTCL to gene signatures of various stages of canine thymocyte and T-cell development. (**A**) Most CD4⁺ PTCLs in Cohort 1 are enriched for the gene signatures of immature canine thymocytes and negatively enriched for the gene signatures of naïve and activated nodal T cells from our canine transcriptomic atlas, while a smaller subset of 5 CD4⁺ PTCLs show low to negative enrichment for all clusters except SP Thymocytes and Proliferating DP Thymocytes. (**B**) In Cohort 2, there is a subset of CD4⁺ PTCL exhibiting enrichment for both immature canine thymocyte gene signatures and naïve and activated T-cell signatures, while the other subset of CD4⁺ PTCL displays only strong enrichment for proliferating DP thymocytes and no other immature thymocyte or mature T-cell signatures from our atlas. Control CD4⁺ lymphocytes from Cohort 2 are enriched for naïve and activated T-cell signatures and negatively enriched for all thymocyte signatures, while control CD4⁺ thymocytes exhibit the strongest enrichment for DP thymocyte and naïve T-cell signatures from our transcriptomic atlas.

In addition to comparing similarities in global gene expression profiles between CD4⁺ PTCL and clusters of our atlas, we also evaluated the expression in CD4⁺ PTCL of individual genes determined by pseudotemporal trajectory analysis to turn “on” and “off” at key points in canine T-cell development. Several genes expressed at early points in thymocyte development and turned off

at later stages of T-cell maturation were significantly overexpressed in canine CD4⁺ PTCL, including *ALPL, BAHCC1, MRC1, PAK3, PTPRF, CD1C, DHDDS, FSTL4, MAML3,* and *NEDD4L* (Figure 44A-B, Table 8).



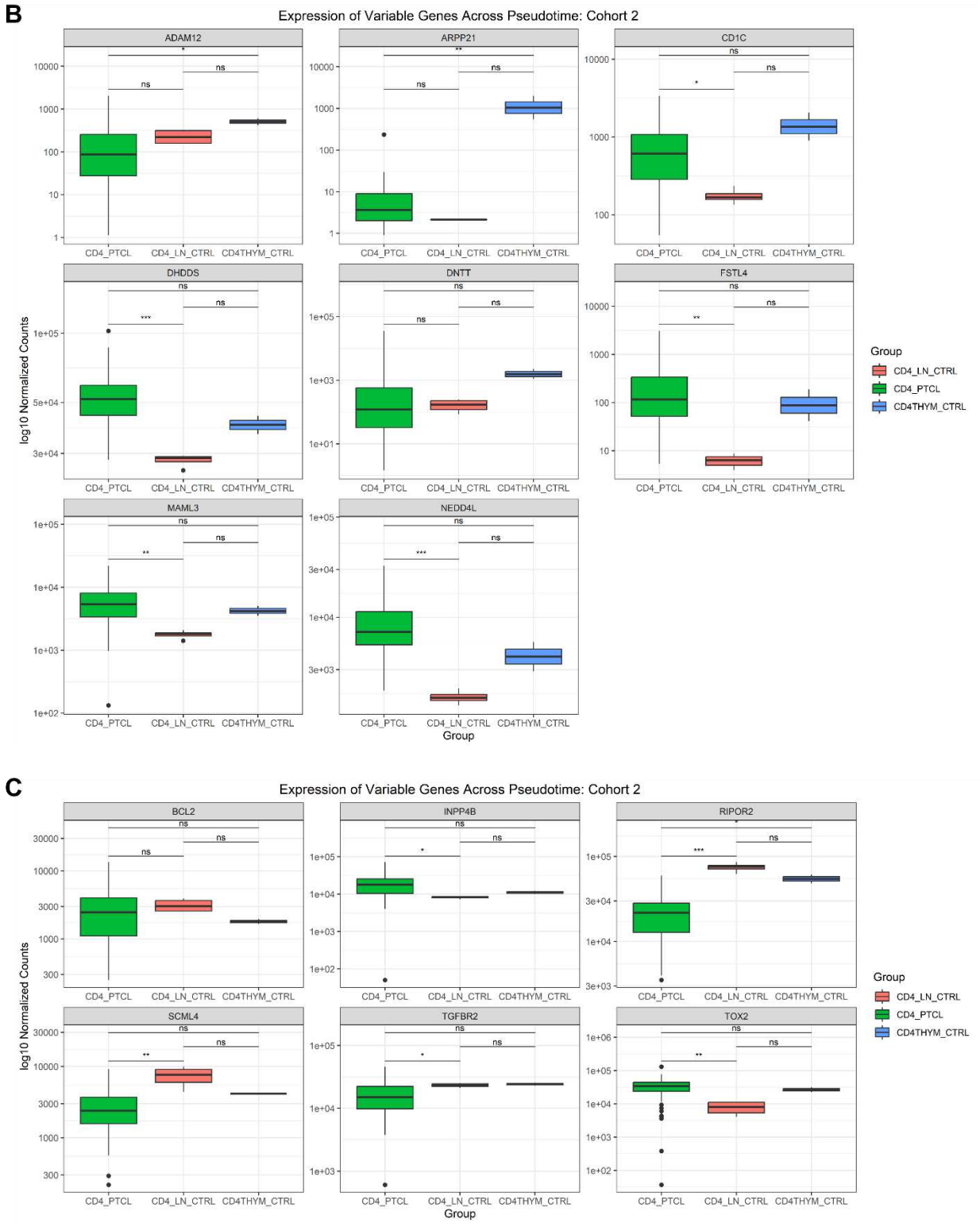
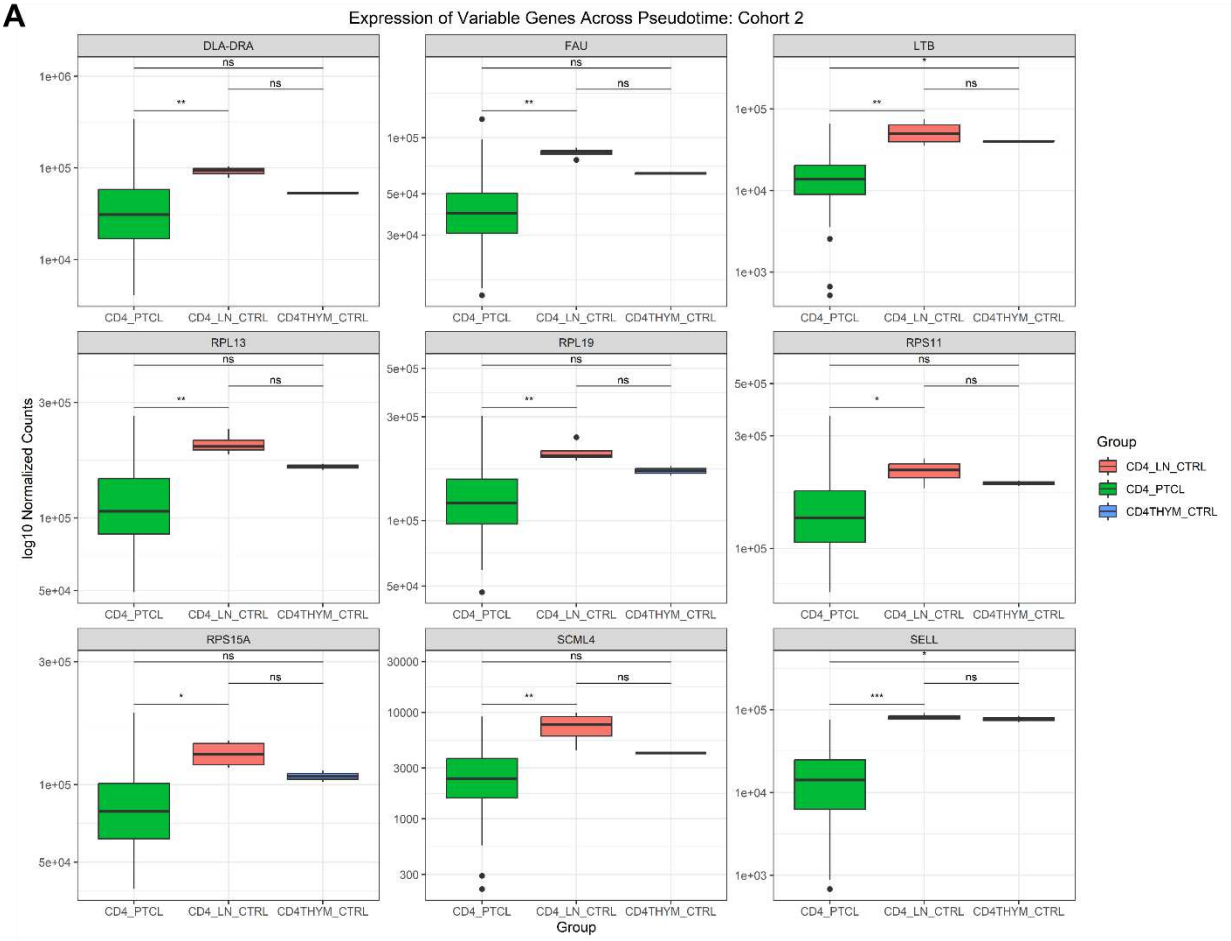


Figure 44: CD4⁺ PTCL overexpresses genes turned on early in T-cell development. Wilcoxon signed-rank tests comparing the means of normalized count data between canine CD4⁺ PTCL, control nodal CD4⁺ lymphocytes, and control CD4⁺ thymocytes. (**A**) Several genes expressed only at the earliest points of thymocyte development are significantly overexpressed by CD4⁺ PTCL compared to control nodal CD4⁺ lymphocytes, including *ALPL, BAHCC1, MRC1, PAK3,* and *PTPRF.* Expression of these genes was not significantly different between CD4⁺ PTCL and control CD4⁺

thymocytes. These results were corroborated with DESeq2 differential expression analysis, which also found that the increased *FBN2* expression in CD4⁺ PTCL compared to control CD4⁺ nodal lymphocytes reached statistical significance (Table 8). (**B**) Genes expressed at the DN and DP thymocyte stages that are also upregulated in CD4⁺ PTCL include *CD1C, DHDDS, FSTL4, MAML3,* and *NEDD4L*. Expression of these genes is not significantly different between CD4⁺ PTCL and control CD4⁺ thymocytes. These results were corroborated with DESeq2 differential expression analysis (Table 8). (**C**) Some genes turned on at the SP thymocyte stages are significantly downregulated in CD4⁺ PTCL compared to control nodal CD4⁺ lymphocytes, including *RIPOR2, SCML4,* and *TGFBR2*, while others are overexpressed, such as *INPP4B* and *TOX2*. Expression of *RIPOR2* was also significantly higher in control CD4⁺ thymocytes compared to CD4⁺ PTCL. However, the decreased expression of *TGFBR2* by CD4⁺ PTCL compared to control CD4⁺ lymphocytes and decreased expression of *RIPOR2* compared to control CD4⁺ thymocytes did not reach statistical significance on differential expression analysis with DESeq2 (Table 8).

On the other hand, genes not turned on until later in T-cell development, such as *FAU, LTB, RPL13, RPL19, RPS15A, SCML4, SELL, B2M, CBLB, GIMAP4, MAF,* and *PARP8*, tended to be significantly downregulated in canine CD4⁺ PTCL (Figure 45, Table 8).



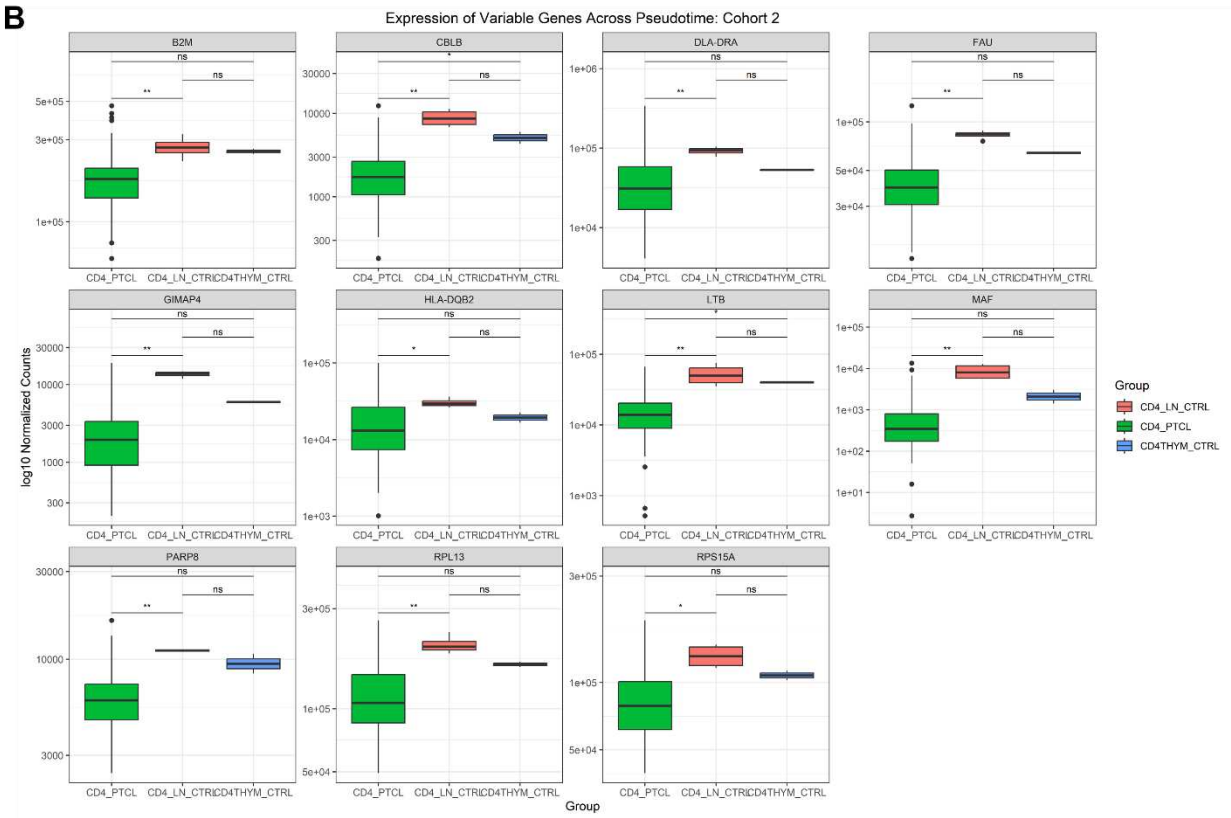


Figure 45: Genes turned on late in T-cell development are downregulated by CD4⁺ PTCL. Wilcoxon signed-rank tests comparing the means of normalized count data between canine CD4⁺ PTCL, control nodal CD4⁺ lymphocytes, and control CD4⁺ thymocytes. (**A**) Several genes turned on at the late SP to early naïve peripheral T cell stages of development (**A**) and mature, activated T-cell subsets (**B**) in our single-cell atlas were significantly downregulated in CD4⁺ PTCL, including *DLA-DRA, FAU, LTB, RPL13, RPL19, RPS11, RPS15A, SCML4, SELL, B2M, CBLB, GIMAP4, HLA-DQB2, MAF,* and *PARP8*. However, the downregulation of *DLA-DRA, RPS11,* and *HLA-DQB2* did not reach statistical significance with DESeq2 differential expression analysis (Table 8). Expression of these genes was not significantly different between CD4⁺ PTCL and control CD4⁺ thymocytes based on DESeq2 differential expression analysis.

Some genes turned on around the SP thymocyte stage of development, including *RIPOR2*, *SCML4,* and *TGFBR2,* were significantly downregulated in CD4⁺ PTCL, while others, such as *INPP4B* and *TOX2,* were upregulated in CD4⁺ PTCL (Figure 44C). With the exception of *ARPP21*, none of these genes were significantly differentially expressed between CD4⁺ PTCL and control CD4⁺ thymocytes based on DESeq2 differential expression analysis (Table 8). Expression of *ARPP21*, a thymocyte-specific RNA-binding protein that promotes *RAG1* expression,[64] was significantly lower in CD4⁺ PTCL compared to control CD4⁺ thymocytes in our bulk RNA-seq study (Table 8). In our single-cell trajectory analysis, *ARPP21* expression was highest in canine DN and DP thymocyte populations.

Table 8: Results of differential expression analysis of changing genes across canine thymocyte and T-cell development between canine CD4⁺ PTCL, control nodal CD4⁺ T cells, and control CD4⁺ thymocytes.

Gene	Expression in canine T-cell development	CD4⁺ PTCL vs. control nodal CD4⁺ T cells		CD4⁺ PTCL vs. control CD4⁺ thymocytes	
		Log2fc	padj	Log2fc	padj
ALPL	Early thymic	1.90	0.0006	-1.18	0.24
BAHCC1	Early thymic	4.23	4.81E-10	0.47	0.89
FBN2	Early thymic	2.98	0.002	-2.68	0.1
MRC1	Early thymic	2.84	6.42E-05	0.22	0.88
PAK3	Early thymic	3.06	0.0002	0.49	0.90
PTPRF	Early thymic	3.61	3.50E-07	-0.36	0.77
TBC1D9	Early thymic	1.76	0.007	1.28	0.35
ADAM12	Early thymic	-0.11	0.95	-1.24	0.48
ARPP21	Early thymic	0.82	0.62	-6.37	9.31E-05
CD1C	Early thymic	2.12	0.0002	-0.74	0.53
DHDDS	Early thymic	0.92	1.45E-07	0.40	0.29
DNTT	Early thymic	3.06	0.08	0.47	0.8
FSTL4	Early thymic	5.71	2.08E-11	1.63	0.48
MAML3	Early thymic	1.84	0.0001	0.58	0.65
NEDD4L	Early thymic	2.49	2.79E-11	1.02	0.21
BCL2	Late thymic	-0.20	0.75	0.70	0.58
INPP4B	Late thymic	1.30	0.01	0.82	0.46
RIPOR2	Late thymic	-1.57	6.31E-05	-1.23	0.11
SCML4	Late thymic	-1.35	0.004	-0.50	0.68
TGFBR2	Late thymic	-0.45	0.32	-0.51	0.60
TOX2	Late thymic	1.79	0.0003	0.40	0.78
DLA-DRA	Naïve and mature T cells	-0.97	0.11	-0.30	0.86
FAU	Naïve and mature T cells	-0.99	0.0005	-0.59	0.33
LTB	Naïve and mature T cells	-1.53	0.002	-1.22	0.21
RPL13	Naïve and mature T cells	-0.73	0.009	-0.45	0.47
RPL19		-0.69	0.007	-0.40	0.47
RPS11		-0.52	0.07	-0.34	0.60
RPS15A	Naïve and mature T cells	-0.64	0.02	-0.33	0.61
SELL	Naïve and mature T cells	-2.07	0.001	-1.98	0.10
B2M	Mature T cells	-0.56	0.04	-0.49	0.40
CBLB	Mature T cells	-1.95	0.0002	-1.14	0.31
GIMAP4	Mature T cells	-2.22	0.0003	-1.11	0.42
HLA-DQB2	Mature T cells	-0.68	0.28	-0.09	0.97
MAF	Mature T cells	-3.13	0.0002	-1.20	0.60
PARP8	Mature T cells	-0.84	0.002	-0.61	0.29

Taken together, canine CD4⁺ PTCL tended to resemble the gene expression profiles of earlier stages of thymocyte development than later stages, suggesting that these tumors do not

originate from mature nodal T cells, but rather malignant transformation likely occurs in an earlier thymic precursor that subsequently emigrates from the thymus to the peripheral lymph nodes.

Discussion

In this study, we investigated the cell of origin of canine $CD4^+$ PTCL through evaluation of expression of surface markers associated with T-cell maturity and activation, gene expression profiling to identify up- and downregulation of genes associated with T-cell immaturity and maturity, respectively, and gene set enrichment and variation analyses to compare gene expression patterns between canine $CD4^+$ PTCL and various stages of human, murine, and canine thymocyte and T-cell development. To allow comparisons of canine $CD4^+$ PTCL gene expression to gene expression at various stages of canine thymocyte and T-cell development, we utilized single-cell RNA-sequencing and pseudotemporal trajectory analysis to develop a single-cell transcriptomic atlas of normal canine thymus and lymph node and identified key genes that change as cells progress through their various stages of development and differentiation. Through these efforts, we found that canine $CD4^+$ PTCL likely originates from a thymic precursor cell of origin. Given that canine $CD4^+$ PTCL is a molecularly similar disease to human PTCL-NOS, this finding may have translational implications that challenge the more conventionally posited mature T-helper cell of origin for human PTCL-NOS.

The cell of origin of human PTCL-NOS remains nebulous, as previous assumptions that the recently identified TBX21-PTCL and GATA3-PTCL subtypes were derived from their normal Th1 and Th2 immune cell counterparts have been more recently met with cautious skepticism given the demonstrable plasticity of T-helper differentiation *in vitro* and the resistance of mature murine T-cells to neoplastic transformation.[1,2] Additionally, *GATA3* has been observed to function as a proto-oncogene in many T-cell tumor types, including precursor neoplasms and non-T-cell neoplasms,

without concurrent expression of other Th2-specific genes, further challenging the ontological relationship between GATA3-expressing PTCLs and Th2 cells.[3]

Despite peripheral nodal involvement in the majority of cases, we observed several features in canine $CD4^+$ PTCL that suggest a more immature cell of origin than the conventionally speculated mature nodal $CD4^+$ T-helper cell phenotype. First, canine $CD4^+$ PTCL cells exhibited low surface MHC class II expression by flow cytometry and significantly decreased transcriptomic expression of MHC class II-associated genes compared to control canine $CD4^+$ lymphocytes. Expression of MHC class II molecules is a feature of all peripheral T cells in dogs, although its expression in the neonatal thymus and on resting T cells is weaker than on activated T cells.[4,5] Thymocytes collected from the thymuses of normal healthy control dogs in our study demonstrated significantly lower expression of MHC class II by flow cytometry compared to $CD4^+$ T-cells from normal lymph nodes. This could suggest that the low MHC class II expression in canine $CD4^+$ PTCL may be a feature of their relative immaturity.

Canine $CD4^+$ PTCL cells also lacked surface expression of CD25 by flow cytometry, which was corroborated by significantly decreased expression of the *IL2RA* gene compared to control nodal $CD4^+$ T-cells. CD25 is the alpha chain of the IL-2 receptor, which heterodimerizes with the beta and gamma chains to result in high-affinity binding of the receptor to IL-2 and promotion of lymphocyte proliferation and survival.[6] In humans and dogs, CD25 expression is associated with an activated lymphocyte phenotype under normal conditions,[7–9] although it has also been expressed in a number of lymphoid neoplasms in both species.[10–12] In human thymocytes, CD25 is transiently expressed in early intrathymic progenitors and early subsets of double-negative (DN) thymocytes, although later subsets of DN thymocytes, DP thymocytes, and most SP thymocytes (with the exception of a portion of CD4 SP thymocytes that go on to differentiate into regulatory T-cells) lack surface expression of CD25.[13] In our study, canine DP thymocytes lacked surface expression of

CD25, and only a small proportion of canine DN thymocytes were $CD25^+$ by flow cytometry. However, a significantly higher proportion of $CD25^+$ cells were observed in the CD4 SP thymocytes and normal nodal $CD4^+$ T-cell populations. Thus, the low surface CD25 expression observed in canine $CD4^+$ PTCL in our study could indicate a tumorigenesis event somewhere around the time of late DN or DP thymocyte development. This is supported by our transcriptomic atlas of normal canine thymus and lymph node as well, where the strongest enrichment for genes upregulated in $CD4^+$ PTCL was appreciated in a cluster of DP thymocytes.

GSEA utilizing our gene expression data revealed significant negative enrichment of canine $CD4^+$ PTCL for gene signatures associated with TCR signaling and decreased expression of TCR signaling genes, suggesting that TCR signaling may be downregulated (or not yet upregulated) in canine $CD4^+$ PTCL. Taken together, the decreased expression of CD25 and MHC class II-associated genes and downregulation of TCR signaling suggest that $CD4^+$ PTCL does not originate from a mature, activated T cell, despite the upregulation of Th2-associated transcription factor *GATA3*.

Although none of the PTCLs in this study expressed cell surface CD34 by flow cytometry, distinguishing them from T-cell acute lymphoblastic leukemia (T-ALL),[14] several genes associated with immaturity or early T-cell development were upregulated in the transcriptome. $CD4^+$ PTCL samples compared to control canine $CD4^+$ lymphocytes showed increased expression of *KIT*, *DNTT*, *CD34,* and *CCR9*. *KIT* is a proto-oncogene that encodes a transmembrane tyrosine kinase and plays an important role in hematopoietic stem cell (HSC) proliferation, differentiation, and survival, and is subsequently downregulated upon HSC maturation.[15] *DNTT* encodes terminal deoxynucleotidyl transferase (TdT), whose expression is normally restricted to immature lymphocytes in the thymus and bone marrow,[16] but is also a frequent feature of acute lymphoid leukemias.[17] *CD34* is widely utilized as a marker of hematopoietic progenitor cells in dogs.[18] *CCR9* encodes a chemokine receptor whose expression is restricted to T cells in the thymus

or small intestine, and in the thymus, it appears to play a role in T-cell development and migration.[19,20] In a single-cell trajectory analysis study of human thymic development, late DP thymocytes went through a transitional CCR9^{high} stage before diverging into mature CD4$^+$ or CD8$^+$ SP cells.[21] This is consistent with historic studies performed in mice, in which CCR9 was not expressed until the DP stage, and was either gradually downregulated prior to thymic egress in the case of CD4$^+$ SP cells or not downregulated at all in the case of CD8$^+$ SP cells.[19,22] Interestingly, while not described as a typical feature of the gene expression profile of human PTCL-NOS, we did observe similar upregulation of *CD34* in the transcriptome of human PTCL-NOS based on publicly available gene expression data from 68 cases of PTCL-NOS published by Etebari et al.[23] (log2 fold change = 5.15, padj = 1.53×10^{-26}) and 28 cases of PTCL-NOS published by Piccaluga et al.[24] (log2 fold change = 0.46, padj = 5.08×10^{-01}). Additionally, a gene expression profiling study by Helgeland et al.[25] found significantly increased expression of *CD34*, *DNTT*, and *CCR9* in CD4 SP thymocytes compared to circulating CD4$^+$ T-cells in infant and adult peripheral blood, suggesting that the upregulation of these genes in canine CD4$^+$ PTCL could favor a thymic precursor cell of origin over a more mature naïve or activated CD4$^+$ T-cell phenotype.

The gene expression profile of canine CD4$^+$ PTCL in our study was enriched for human and murine gene signatures associated with early T-lymphocytes, intrathymic T progenitor cells, CD4 SP thymocytes, and DP thymocytes compared to naïve circulating CD4$^+$ T-cells. There was no significant enrichment for human or murine Th1, Th2, Th17 or regulatory T-cell signatures in canine CD4$^+$ PTCL compared to control CD4$^+$ lymphocytes. GSVA evaluating the enrichment for these cell types at an individual sample level found that a subset of CD4$^+$ PTCL exhibited strong enrichment for genes upregulated in human and murine immature thymocytes compared to naïve or mature T cells, while other CD4$^+$ PTCLs showed only enrichment for genes expressed by human and murine CD4 SP thymocytes and naïve T cells. GSVA using gene signatures derived from our canine single-

cell transcriptomic atlas of normal canine thymus and lymph node showed a similar disparity, where a subset of $CD4^+$ PTCL exhibited strong enrichment for various immature DN and DP thymocyte gene signatures, while other $CD4^+$ PTCLs were non-enriched or negatively enriched for those earliest thymic cell types and instead demonstrated enrichment only for canine SP thymocyte signatures. This variable enrichment within $CD4^+$ PTCL for different stages of early T-cell development could suggest that malignant transformation may occur at different developmental stages, with some PTCLs arising from more immature DN to DP thymocytes and others arising from SP thymocytes or naïve recent thymic emigrants, or that these neoplastic T cells retain some ability to differentiate, leading to the formation of subtypes reflecting varying degrees of T-cell differentiation.

By performing pseudotemporal trajectory analysis on single-cell RNA-seq data from normal canine thymus and lymph node samples, we were able to identify genes that turn on and off at various stages of thymocyte and T-cell development and evaluate the expression of these genes in $CD4^+$ PTCL to further probe the cell of origin of these tumors. Our trajectory analysis identified not only genes whose biologic roles in T-cell development are well-established, which helped validate the cell type assignments of our single-cell atlas, but also several genes whose function in T-cell maturation is not well established. Compared to control nodal $CD4^+$ T cells, $CD4^+$ PTCL upregulated several genes expressed only in early cortical stages of thymocyte development, variably up- and downregulated genes turned on around the SP thymocyte stage, and downregulated genes turned on in late stages of T-cell development in the periphery.

Some notable genes expressed only at the earliest points of thymocyte development in our trajectory analysis that were also upregulated in $CD4^+$ PTCL include *ALPL, BAHCC1, MAML3,* and *CD1C*. *BAHCC1* recognizes genes marked by trimethylated histone H3 lysine 27 (H3K27me3) to enforce gene silencing.[26] H3K27me3 regulates transcription of thousands of genes, most of which

have been found to be associated with embryonic development and cell fate decisions.[27] In studies of acute lymphoblastic leukemia, *BAHCC1* was found to play a crucial role in maintaining long-term repopulating hematopoietic stem cells through repression of cyclin dependent kinase inhibitor *CDKN1C*.[28] *MAML3* is a coactivator of Notch signaling, a critical signaling pathway for T lineage commitment and thymocyte development and survival.[29] *CD1C* encodes a CD1 glycoprotein, and in other mammalian species, CD1 glycoproteins are expressed in DN thymocytes prior to β-selection and are progressively downregulated as thymocytes continue to mature.[30]

Notable genes turned on at the SP thymocyte stage of our trajectory analysis that were downregulated by $CD4^+$ PTCL include *RIPOR2* and *TGFBR2*, both important in maintaining pre-activated T cells in a quiescent state prior to TCR stimulation.[31,32] Genes turned on at the SP thymocyte stage that were upregulated in $CD4^+$ PTCL included *INPP4B* and *TOX2*. *INPP4B* promotes PI3K/AKT signaling in the context of maintaining innate lymphocyte and NK cell populations in peripheral tissues (including peripheral lymph nodes).[33] *TOX2* is best known for playing a role in the differentiation of NK cells and follicular T-helper cells,[34] but it has also been shown to be upregulated in DP thymocytes following positive selection,[35] although its function at this stage of T-cell development is not well established.

Finally, noteworthy genes turned on late in T-cell development in our trajectory analysis that were significantly downregulated in $CD4^+$ PTCL included *CBLB*, which mitigates the T-cell response by promoting internalization of the TCR in activated T cells;[36] *GIMAP4*, which is expressed in DN thymocytes in response to pre-TCR signaling before being transiently downregulated at the DP stage and re-expressed in SP thymocytes and peripheral T cells;[37] MHC class I gene *B2M*; T-cell homing molecule *SELL* (L-selectin), which is generally most highly expressed by naïve and central memory T cells;[38] *MAF*, a transcription factor highly expressed by terminally differentiated T cells after TCR activation;[39–41] and marker of T-cell activation *LTB*.[42]

Despite the increased expression of markers of immaturity in the transcriptome of canine CD4$^+$ PTCL, enrichment of these tumor cells for early thymocyte gene signatures, and downregulation of TCR signaling genes and other markers of T-cell maturity demonstrated by our study, CD4$^+$ PTCLs still exhibit clonal rearrangement of their TCR, downregulation of *RAG1* and *RAG2* genes (Supplemental Figure 9), and a subset still express mature T-cell surface antigen CD5, features more typical of post-thymic T cells. The reason for these discrepancies is unclear. T-cell malignancies that originate from thymic precursors but still express a TCR have been reported in the human literature. For example, in anaplastic large cell lymphoma (ALCL), TCR rearrangement analyses identified malignant transformation of tumor cells in thymocytes prior to TCR β-rearrangement, but RAG-independent TCR expression driven by the oncogenic NPM-ALK fusion protein still allowed these cells to exit the thymus, and downregulation of the TCR was required for subsequent lymphomagenesis.[43] The negative enrichment of canine CD4$^+$ PTCL for the TCR signaling pathway and upregulation of various markers of immaturity could suggest a similar process in canine CD4$^+$ PTCL, although similar oncogenic gene fusions have yet to be investigated in this disease. Another possibility could include the reversion of mature CD4$^+$ T-cells to a more immature phenotype following malignant transformation. Primary human CD4$^+$ T cells transduced with the previously mentioned *NPM-ALK* fusion have been shown to adopt a more immature gene expression profile and undergo malignant transformation following TCR stimulation with anti-CD3/28.[44] In mice, *PAX5* deletions have been shown to result in reversion of mature B cells to an uncommitted precursor cell phenotype and induce the formation of progenitor cell lymphomas.[45] Additional studies investigating similar driver mutations in canine PTCL may help identify mechanisms contributing to the contradictory immature tumor cell transcriptome with surface expression of mature T-cell markers.

Conclusions

Our study suggests that canine $CD4^+$ PTCL may arise from a thymic precursor cell of origin, based on the observed enrichment for gene signatures associated with early thymocyte progenitors; downregulation of TCR signaling; decreased expression of CD25 (*IL2RA*), MHC class II-associated genes, and other markers of T-cell maturity; and increased expression of markers of immaturity. Our findings have important implications for the pathogenesis of canine PTCL, as this precursor-like gene expression profile with downregulation of TCR signaling could suggest that chronic inflammation leading to prolonged TCR stimulation is likely not implicated as a predisposing factor for the development of canine PTCL. A thymic precursor cell of origin may also have therapeutic implications for these tumors, as pathways involved in thymocyte survival, like Notch signaling, may represent an opportunity for more targeted therapies. *In vitro* and *in vivo* studies investigating the dependency of these tumor cells on these pathways are an important future direction both to further validate our proposed cell of origin and explore the potential therapeutic opportunities illuminated by this cell of origin to improve outcomes in dogs with $CD4^+$ PTCL.

REFERENCES

1. Hirota, K. *et al.* Fate mapping of IL-17-producing T cells in inflammatory responses. *Nat. Immunol.* 12, 255–263 (2011).

2. Zhou, X. *et al.* Instability of the transcription factor Foxp3 leads to the generation of pathogenic memory T cells in vivo. *Nat. Immunol.* 10, 1000–1007 (2009).

3. Mamand, S., Carr, M., Allchin, R. L., Ahearne, M. J. & Wagner, S. D. Interleukin-2-inducible T-cell kinase inhibitors modify functional polarization of human peripheral T-cell lymphoma cells. *Blood Adv.* 3, 705–710 (2019).

4. de Leval, L. & Gaulard, P. Cellular origin of T-cell lymphomas. *Blood* 123, 2909–2910 (2014).

5. Newrzela, S. *et al.* Resistance of mature T cells to oncogene transformation. *Blood* 112, 2278–2286 (2008).

6. Geng, X. *et al.* GATA-3 is a proto-oncogene in T-cell lymphoproliferative neoplasms. *Blood Cancer J.* 12, 149 (2022).

7. Harris, L. J. *et al.* Canine CD4+ T-cell lymphoma identified by flow cytometry exhibits a consistent histomorphology and gene expression profile. *Vet. Comp. Oncol.* 17, 253–264 (2019).

8. Hughes, K. *et al.* Canine T zone lymphoma is a tumor of mature, previously activated αβ T cells. *Vet. Immunol. Immunopathol.* 269, 110725 (2024).

9. Doveren, R. F., van der Linden, C. J., Spronken, E. E., Groenewegen, G. & Buurman, W. A. Canine MHC-class II antigens on B and T lymphocytes. *Tissue Antigens* 27, 87–98 (1986).

10. Doveren, R. F., Buurman, W. A., Schutte, B., Groenewegen, G. & van der Linden, C. J. Class II antigens on canine T lymphocytes. *Tissue Antigens* 25, 255–265 (1985).

11. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

12. Kassambara, A. *Ggpubr: 'ggplot2' Based Publication Ready Plots*. (2023).

13. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS J. Integr. Biol.* 16, 284–287 (2012).

14. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740 (2011).

15. Helgeland, H. *et al.* Transcriptome profiling of human thymic CD4+ and CD8+ T cells compared to primary peripheral T cells. *BMC Genomics* 21, 350 (2020).

16. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics* 14, 7 (2013).

17. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43, e47–e47 (2015).

18. Warnes, G. *et al.* gplots: Various R Programming Tools for Plotting Data. *R package version 3.1.3* https://CRAN.R-project.org/package=gplots (2022).

19. Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Res.* 48, D682–D688 (2020).

20. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2022).

21. Park, J.-E. *et al.* A cell atlas of human thymic development defines T cell repertoire formation. *Science* 367, eaay3224 (2020).

22. Domínguez Conde, C. *et al.* Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science* 376, eabl5197.

23. Ammons, D. T. *et al.* A single-cell RNA sequencing atlas of circulating leukocytes from healthy and osteosarcoma affected dogs. *Front. Immunol.* 14, (2023).

24. Rouillard, A. D. *et al.* The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. *Database* 2016, baw100 (2016).

25. Martens, M. *et al.* WikiPathways: connecting communities. *Nucleic Acids Res.* 49, D613–D621 (2021).

26. Lee, M. S., Hanspers, K., Barker, C. S., Korn, A. P. & McCune, J. M. Gene expression profiles during human CD4+ T cell differentiation. *Int. Immunol.* 16, 1109–1124 (2004).

27. Szabo, P. A. *et al.* Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nat. Commun.* 10, 4706 (2019).

28. Gramaglia, I., Mauri, D. N., Miner, K. T., Ware, C. F. & Croft, M. Lymphotoxin alphabeta is expressed on recently activated naive and Th1-like CD4 cells but is down-regulated by IL-4 during Th2 differentiation. *J. Immunol. Baltim. Md 1950* 162, 1333–1338 (1999).

29. Yu, J. & Liu, S.-L. Emerging Role of LY6E in Virus-Host Interactions. *Viruses* 11, (2019).

30. Klein, F. *et al.* Dntt expression reveals developmental hierarchy and lineage specification of hematopoietic progenitors. *Nat. Immunol.* 23, 505–517 (2022).

31. Choi, S. M., Park, H. J., Choi, E. A., Jung, K. C. & Lee, J. I. CD1b glycoprotein, a crucial marker of thymocyte development during T cell maturation in cynomolgus monkeys. *Sci. Rep.* 13, 14388 (2023).

32. Li, K.-S. *et al.* NT5DC2 promotes tumor cell proliferation by stabilizing EGFR in hepatocellular carcinoma. *Cell Death Dis.* 11, 335 (2020).

33. Fu, G. *et al.* Themis controls thymocyte selection through regulation of T cell antigen receptor–mediated signaling. *Nat. Immunol.* 10, 848–856 (2009).

34. Blonska, M. *et al.* Activation of the Transcription Factor c-Maf in T Cells Is Dependent on the CARMA1-IKKβ Signaling Cascade. *Sci. Signal.* 6, ra110–ra110 (2013).

35. Yang, Y., Ochando, J., Yopp, A., Bromberg, J. S. & Ding, Y. IL-6 Plays a Unique Role in Initiating c-Maf Expression during Early Stage of CD4 T Cell Activation1. *J. Immunol.* 174, 2720–2729 (2005).

36. Andris, F. *et al.* The Transcription Factor c-Maf Promotes the Differentiation of Follicular Helper T Cells. *Front. Immunol.* 8, (2017).

37. Galkowska, H., Waldemar, L. O. & Wojewodzka, U. Reactivity of antibodies directed against human antigens with surface markers on canine leukocytes. *Vet. Immunol. Immunopathol.* 53, 329–334 (1996).

38. Helfand, S. C., Modiano, J. F. & Nowell, P. C. Immunophysiological studies of interleukin-2 and canine lymphocytes. *Vet. Immunol. Immunopathol.* 33, 1–16 (1992).

39. Reddy, M., Eirikis, E., Davis, C., Davis, H. M. & Prabhakar, U. Comparative analysis of lymphocyte activation marker expression and cytokine secretion profile in stimulated human peripheral blood mononuclear cell cultures: an in vitro model to monitor cellular immune function. *J. Immunol. Methods* 293, 127–142 (2004).

40. Dong, C. *et al.* ICOS co-stimulatory receptor is essential for T-cell activation and function. *Nature* 409, 97–101 (2001).

41. Kerkhoff, C., Klempt, M. & Sorg, C. Novel insights into structure and function of MRP8 (S100A8) and MRP14 (S100A9). *Biochim. Biophys. Acta BBA - Mol. Cell Res.* 1448, 200–211 (1998).

42. Averill, M. M., Kerkhoff, C. & Bornfeldt, K. E. S100A8 and S100A9 in Cardiovascular Biology and Disease. *Arterioscler. Thromb. Vasc. Biol.* 32, 223–229 (2012).

43. Li, A. *et al.* TOX2 nuclear-cytosol translocation is linked to leukemogenesis of acute T-cell leukemia by repressing TIM3 transcription. *Cell Death Differ.* 31, 1506–1518 (2024).

44. Mingueneau, M. *et al.* The transcriptional landscape of αβ T cell differentiation. *Nat. Immunol.* 14, 619–632 (2013).

45. Chopp, L., Redmond, C., O'Shea, J. J. & Schwartz, D. M. From thymus to tissues and tumors: A review of T-cell biology. *J. Allergy Clin. Immunol.* 151, 81–97 (2023).

46. Sun, L., Su, Y., Jiao, A., Wang, X. & Zhang, B. T cells in health and disease. *Signal Transduct. Target. Ther.* 8, 235 (2023).

47. Mallory, M. J. *et al.* Induced transcription and stability of CELF2 mRNA drives widespread alternative splicing during T-cell signaling. *Proc. Natl. Acad. Sci.* 112, E2139–E2148 (2015).

48. Goldstein, I. *et al.* Expression of the α1β1 integrin, VLA-1, marks a distinct subset of human CD4+ memory T cells. *J. Clin. Invest.* 112, 1444–1454 (2003).

49. Froehlich, J. *et al.* FAM65B controls the proliferation of transformed and primary T cells. *Oncotarget* 7, 63215–63225 (2016).

50. Gavin, M. A. *et al.* Foxp3-dependent programme of regulatory T-cell differentiation. *Nature* 445, 771–775 (2007).

51. Tu, E. *et al.* T Cell Receptor-Regulated TGF-β Type I Receptor Expression Determines T Cell Quiescence and Activation. *Immunity* 48, 745-759.e6 (2018).

52. Feng, M. *et al.* Fatty Acids Support the Fitness and Functionality of Tumor-Resident CD8+ T Cells by Maintaining SCML4 Expression. *Cancer Res.* 83, 3368–3384 (2023).

53. Lee, M. G., Sharrow, S. O., Farr, A. G., Singer, A. & Udey, M. C. Expression of the homotypic adhesion molecule E-cadherin by immature murine thymocytes and thymic epithelial cells. *J. Immunol.* 152, 5653–5659 (1994).

54. Kutleša, S. *et al.* E-cadherin-mediated interactions of thymic epithelial cells with CD103+ thymocytes lead to enhanced thymocyte cell proliferation. *J. Cell Sci.* 115, 4505–4515 (2002).

55. Umetsu, S. E. *et al.* TIM-1 induces T cell activation and inhibits the development of peripheral tolerance. *Nat. Immunol.* 6, 447–454 (2005).

56. Zhai, Y. *et al.* Opposing regulatory functions of the TIM3 (HAVCR2) signalosome in primary effector T cells as revealed by quantitative interactomics. *Cell. Mol. Immunol.* 18, 1581–1583 (2021).

57. Gratiot-Deans, J., Merino, R., Nuñez, G. & Turka, L. A. Bcl-2 expression during T-cell development: early loss and late return occur at specific stages of commitment to differentiation and survival. *Proc. Natl. Acad. Sci. U. S. A.* 91, 10685–10689 (1994).

58. Grayson, J. M., Zajac, A. J., Altman, J. D. & Ahmed, R. Cutting Edge: Increased Expression of Bcl-2 in Antigen-Specific Memory CD8+ T Cells1. *J. Immunol.* 164, 3950–3954 (2000).

59. Feng, H. *et al.* T-Lymphoblastic Lymphoma Cells Express High Levels of BCL2, S1P1, and ICAM1, Leading to a Blockade of Tumor Cell Intravasation. *Cancer Cell* 18, 353–366 (2010).

60. Tanaka, S. *et al.* Sox5 and c-Maf cooperatively induce Th17 cell differentiation via RORγt induction as downstream targets of Stat3. *J. Exp. Med.* 211, 1857–1874 (2014).

61. Tian, Y. *et al.* SOX-5 activates a novel RORγt enhancer to facilitate experimental autoimmune encephalomyelitis by promoting Th17 cell differentiation. *Nat. Commun.* 12, 481 (2021).

62. Schnell, S., Démollière, C., van den Berk, P. & Jacobs, H. Gimap4 accelerates T-cell death. *Blood* 108, 591–599 (2006).

63. Peng, V. *et al.* Inositol phosphatase INPP4B sustains ILC1s and intratumoral NK cells through an AKT-driven pathway. *J. Exp. Med.* 221, e20230124 (2024).

64. Xu, M. *et al.* The thymocyte-specific RNA-binding protein Arpp21 provides TCR repertoire diversity by binding to the 3'-UTR and promoting Rag1 mRNA expression. *Nat. Commun.* 15, 2194 (2024).

65. Lowenthal, J. W., Zubler, R. H., Nabholz, M. & MacDonald, H. R. Similarities between interleukin-2 receptor number and affinity on activated B and T lymphocytes. *Nature* 315, 669–672 (1985).

66. Seelig, D. M. *et al.* Canine T-zone lymphoma: unique immunophenotypic features, outcome, and population characteristics. *J. Vet. Intern. Med.* 28, 878–886 (2014).

67. Mizutani, N. *et al.* Evaluation of CD25-positive cells in relation to the subtypes and prognoses in various lymphoid tumours in dogs. *Vet. Immunol. Immunopathol.* 173, 39–43 (2016).

68. Nakase, K. *et al.* Differential expression of interleukin-2 receptors (α and β chain) in mature lymphoid neoplasms. *Am. J. Hematol.* 46, 179–183 (1994).

69. Ceredig, R. & Rolink, T. A positive look at double-negative thymocytes. *Nat. Rev. Immunol.* 2, 888–897 (2002).

70. Arber, D. A. *et al.* The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 127, 2391–2405 (2016).

71. Edling, C. E. & Hallberg, B. c-Kit—A hematopoietic cell essential receptor tyrosine kinase. *Int. J. Biochem. Cell Biol.* 39, 1995–1998 (2007).

72. Bollum, F. Terminal deoxynucleotidyl transferase as a hematopoietic cell marker. *Blood* 54, 1203–1215 (1979).

73. Kaleem, Z. *et al.* Flow Cytometric Analysis of Acute Leukemias: Diagnostic Utility and Critical Analysis of Data. *Arch. Pathol. Lab. Med.* 127, 42–48 (2003).

74. McSweeney, P. A. *et al.* Characterization of Monoclonal Antibodies That Recognize Canine CD34. *Blood* 91, 1977–1986 (1998).

75. Uehara, S., Grinberg, A., Farber, J. M. & Love, P. E. A Role for CCR9 in T Lymphocyte Development and Migration1. *J. Immunol.* 168, 2811–2819 (2002).

76. Uehara, S., Song, K., Farber, J. M. & Love, P. E. Characterization of CCR9 Expression and CCL25/Thymus-Expressed Chemokine Responsiveness During T Cell Development: CD3highCD69+ Thymocytes and γδTCR+ Thymocytes Preferentially Respond to CCL251. *J. Immunol.* 168, 134–142 (2002).

77. Etebari, M. *et al.* Transcriptional Analysis of Lennert Lymphoma Reveals a Unique Profile and Identifies Novel Therapeutic Targets. *Front. Genet.* 10, (2019).

78. Piccaluga, P. P. *et al.* Gene expression analysis of peripheral T cell lymphoma, unspecified, reveals distinct profiles and new potential therapeutic targets. *J. Clin. Invest.* 117, 823–834 (2007).

79. Malcolm, T. I. M. *et al.* Anaplastic large cell lymphoma arises in thymocytes and requires transient TCR expression for thymic egress. *Nat. Commun.* 7, 10087 (2016).

80. Pawlicki, J. M. *et al.* NPM–ALK-Induced Reprogramming of Mature TCR-Stimulated T Cells Results in Dedifferentiation and Malignant Transformation. *Cancer Res.* 81, 3241–3254 (2021).

81. Cobaleda, C., Jochum, W. & Busslinger, M. Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. *Nature* 449, 473–477 (2007).

CHAPTER THREE: Recurrent fusion genes and other structural variants in canine CD4$^+$ peripheral T-cell lymphoma

Summary

Fusion genes are important oncogenic driver mutations in several hematopoietic malignancies, including human peripheral T-cell lymphoma (PTCL). In addition to providing insight into the pathogenesis of these tumors, these mutations can also represent an opportunity for targeted therapies. While several recurrent fusion genes have been described in human PTCL, this phenomenon has not been described in canine PTCL.

Chromosomal aberrations leading to the formation of fusion genes, such as translocations and large insertions and deletions (INDELs), are large (>1 kilobase) DNA rearrangements known as structural variants. While short variants like single nucleotide polymorphisms (SNPs) and small INDELs can be detected with short-read sequencing methods like whole genome/exome sequencing and RNA-sequencing (RNA-seq), detection of large structural variants has historically required conventional metaphase karyotyping or targeted molecular assays like fluorescence *in situ* hybridization (FISH). Recently, long-read sequencing has risen to prominence as a new methodology for detection of structural variants in high molecular weight DNA samples.

In Chapter One, we established that canine CD4$^+$ PTCL and human PTCL-not otherwise specified (PTCL-NOS) are molecularly similar diseases. The discovery of fusion genes or other structural variants driving canine CD4$^+$ PTCL may therefore offer insight into the pathogenesis of and alternative treatment options for this malignancy in both humans and dogs. In this study, we sought to identify and characterize recurrent fusion genes in canine CD4$^+$ PTCL and utilize gene expression profiling to predict the potential impacts of these mutations on the pathogenesis of CD4$^+$ PTCL.

To accomplish this, fusion gene candidates were identified from bulk RNA-seq experiments of canine CD4$^+$ PTCL using the fusion caller algorithms FusionCatcher or STAR-Fusion. Fusion gene candidates that were called in multiple dogs, or that involved at least one cancer-associated gene partner, were selected for validation by reverse transcription polymerase chain reaction (RT-PCR) and Sanger sequencing. Single-molecule real-time (SMRT) long-read sequencing was performed on a subset of these dogs to determine if fusion gene mRNA transcripts had a corresponding chromosomal aberration, and to identify other large genomic structural variants in canine CD4$^+$ PTCL.

Approximately 3,000 total fusion gene candidates were predicted by fusion callers from bulk RNA-seq data of CD4$^+$ PTCL, and 27 candidate fusion genes that were recurrent or that involved at least one cancer-associated gene were selected for further study. 108 canine CD4$^+$ PTCL samples and 18 samples from control dogs without lymphoma were screened by RT-PCR for these 27 fusions. 10 of these 27 fusions were identified in multiple samples of CD4$^+$ PTCL and samples from control dogs without lymphoma. 3 of these 27 fusions were tumor-specific, identified in multiple samples of CD4$^+$ PTCL but not in samples from control dogs without lymphoma. Some of these fusion genes were associated with changes in enrichment of tumors for certain intracellular signaling pathways or elements of the tumor microenvironment, which may offer insight into potential mechanisms of their contribution to CD4$^+$ PTCL pathogenesis and options for targeted therapies.

Long-read sequencing identified over 120,000 structural variants in CD4$^+$ PTCL samples, with 2% of these being chromosomal breaks that could result in fusion genes. While none of the fusion gene candidates validated by PCR in CD4$^+$ PTCL had a corresponding chromosomal aberration identified by long-read sequencing, 3 fusion genes not included as part of our PCR assay were identified both at the DNA level by long-read sequencing and at the transcript level by bulk

RNA-seq. Long-read sequencing additionally identified structural variants in several genes commonly mutated in human PTCL-NOS, including *PTEN, TP53, CDKN2A, STAT3,* and *BCL11B*.

These findings support that, similar to human PTCL, recurrent fusion genes are a feature of canine $CD4^+$ PTCL that may arise from a combination of chromosomal aberrations and alternative RNA splicing events, and gene expression profiling comparing fusion-positive to fusion-negative $CD4^+$ PTCLs uncovered potential mechanisms by which some fusions may be contributing to the pathogenesis of canine $CD4^+$ PTCL that warrant further investigation. In addition to fusion genes, structural variants in genes commonly mutated in human PTCL-NOS are a feature of canine $CD4^+$ PTCL, further supporting the molecular similarity of these diseases. The finding of fusion gene transcripts in normal dogs highlights the necessity of rigorous validation of fusion gene candidates predicted from high-throughput sequencing techniques.

Introduction

An overview of fusion genes in PTCL-NOS

Cancer is a disease of the genome, and the identification of mutations that drive its onset and progression is a critical step in improving the ability to predict, prevent, and treat these malignancies. In Chapter One, we discussed the impact of small variants, such as single nucleotide polymorphisms (SNPs) or small insertions and deletions (INDELs), which can disrupt normal protein translation to deleterious effect. In the current study, we sought to understand the role of larger structural variants in the pathogenesis of canine $CD4^+$ PTCL, with a particular emphasis on the discovery of recurrent gene fusions.

Fusion genes are a type of mutation in which two previously independent genes become juxtaposed, transcribed, and potentially translated together.[1,2] Given the inherent genomic instability of cancer cells, gene fusions happen frequently, with the majority expected to be passenger

mutations with no functional relevance.[1] However, some fusion genes can contribute to oncogenesis by two major mechanisms: altered regulation of gene expression, or translation of novel fusion proteins. Deregulation can include fusion of the regulatory elements of one gene to the complete coding sequence of another gene, resulting in altered expression of that gene, such as when a promoter is fused to an oncogene in a process known as *promoter exchange*.[1] Alternatively, fusion genes may form novel fusion proteins if the coding regions of each partner fuse in a way that creates an in-frame sequence, and these fusion proteins may then exert an oncogenic effect in the cell.[1,3]

Several recurrent fusion genes have been described in human PTCL-NOS. The *ITK-SYK* fusion gene resulting from a t(5;9)(q33;q22) translocation has been reported in up to 17% of PTCL-NOS.[4] *ITK-SYK* encodes a fusion tyrosine kinase that constitutively associates with lipid rafts and triggers activation of proximal TCR signaling proteins in the absence of antigen, resulting in constitutively activated TCR signaling and oncogenic transformation.[5] Expression of this fusion gene in the bone marrow of mice was sufficient to induce T-cell lymphoproliferative disease, supporting its role as a driver mutation for T-cell malignancy.[6] *FYN-TRAF3IP2* and *KHDRBS1-LCK* are two fusions of subsets of human PTCL-NOS that constitutively activate NF-κB signaling downstream of the TCR to drive tumorigenesis.[7,8] Fusions involving *VAV1* have been reported in up to 11% of PTCL-NOS and result in loss of *VAV1*'s regulatory domains and increased downstream *RAC1* activation, which drives oncogenesis through a variety of molecular mechanisms.[9]

In addition to the insight they offer into tumorigenesis, fusion genes also represent a promising target for therapeutic intervention. The development of kinase inhibitors like imatinib that target the BCR-ABL fusion protein revolutionized the treatment of chronic myeloid leukemia, where the *BCR-ABL* fusion gene is present in over 95% of patients.[10] In PTCL-NOS, inhibitors of SYK and inhibitors of IκB kinase, a master regulator of the NF-κB signaling pathway, have been

effective at preventing the development of T-cell lymphoproliferative disease in *ITK-SYK* expressing mice or reducing tumor burden in mice with the *FYN-TRAF3IP2* fusion, respectively.[6,8]

Mechanisms of fusion gene synthesis

Fusion genes can originate at the DNA or RNA level. At the DNA level, they are driven by large chromosomal aberrations like insertions, deletions, translocations, inversions, tandem duplications, and chromothripsis.[2] The first fusion gene discovered, *BCR-ABL*, results from a reciprocal t(9;22)(q34;q11) chromosomal translocation that fuses the *BCR* and *ABL1* genes, resulting in their translation and transcription as a hybrid fusion protein, BCR-ABL, that has constitutive tyrosine kinase activity.[1,10] Intra-chromosomal rearrangements, like inversions or large deletions, can also cause a fusion event at the DNA level. The *STIL-TAL1* fusion gene found in 25% of human T-cell acute lymphoblastic leukemia (T-ALL) results from an interstitial deletion on chromosome 1 and places the *TAL1* oncogene under the regulatory control of the *STIL* gene promoter, prompting its overexpression and contribution to leukemogenesis.[11] The *EML4-ALK* fusion gene in non-small cell lung cancer results from a (p21p23) inversion and reversal of one of the genes on chromosome 2.[12] The resulting EML4-ALK fusion protein allows direct dimerization of ALK independent of ligand binding, constitutively activating RAS/ERK, STAT3, and mTOR signaling to tumorigenic effect.[12]

At the RNA level, fusion genes can arise as a result of alternative splicing events where exons from more than one gene are joined together. RNA splicing is a normal physiologic mechanism by which introns—the noncoding segments of DNA important for the regulation of gene expression—are excised from a pre-mRNA molecule to produce a mature mRNA composed only of exons that contain protein-coding information.[2] Fusions arising from cis-splicing occur when exons of adjacent genes located on the same strand (also known as “readthrough fusions”) are joined, while trans-splicing fuses exons from two separate RNA transcripts together.[1,2] Although most fusions resulting from cis-splicing likely represent random events with minimal to no effect on cell

function, some have been found to be expressed in an apparently non-random, tissue-specific manner and have been associated with the development or progression of neoplasia.[1] For example, expression of the *SLC45A3-ELK4* fusion, which is formed by cis-splicing without an associated chromosomal aberration, is positively associated with the progression of prostate cancer in humans, and silencing of the fusion RNA (but not *ELK4* alone) inhibits proliferation of these cells *in vitro*.[13]

One example of an oncogenic gene fusion resulting from trans-splicing is the *CYCLIN D1-TROP2* RNA fusion, which has been detected in a variety of human cancer cells, including gastrointestinal, ovarian, mammary, and endometrial cancer cells, and shown to be capable of inducing malignant transformation *in vitro* and *in vivo*.[14] *CYCLIN D1* and *TROP2* originate on different chromosomes, but no evidence of chromosomal translocation accompanies the *CYCLIN D1-TROP2* fusion RNA transcript in these tumor cells. This post-transcriptional fusion results in loss of the instability sequences in the *CYCLIN D1* 3' untranslated region (UTR) that typically regulate its half-life, resulting in its inappropriate expression and subsequent neoplastic transformation of cells. The *TROP2* sequences appear to contribute a dominant stabilizing effect, as the fusion mRNA results in greater stabilization than simple removal of the *CYCLIN D1* 3' UTR sequences alone.[14]

Interestingly, fusion mRNA transcripts resulting from trans-splicing events are commonly detectable in normal as well as neoplastic tissues.[15] In some cases, identical fusion transcripts will only be accompanied by a chromosomal translocation in cases of malignant transformation, while non-diseased tissue will express only the fusion transcript without a concurrent chromosomal aberration. The most notable examples of this include the *JAZF1-SUZ12* and *PAX3-FOXO1* fusions. *JAZF1-SUZ12* results from a recurrent t(7;17)(p15;q21) chromosomal translocation in endometrial stromal tumors, but the identical *JAZF1-SUZ12* fusion RNA transcript is still detectable in normal endometrial cells without the chromosomal translocation.[16] A chromosomal translocation resulting in the *PAX3-FOXO1* fusion gene is only detectable in rhabdomyosarcoma, but identical *PAX3-FOXO1*

fusion RNA produced by trans-splicing is transiently expressed during normal skeletal muscle cell differentiation.[17] Studies by Gupta et al.[18,19] prompted by these observations discovered that transfection of *JAZF1-SUZ12* fusion RNA and *TMPRSS2-ETV1* fusion RNA (a fusion present in ~50% of human prostate cancers) in fusion-negative human endometrial stromal cells and prostate cancer cells, respectively, was capable of inducing the associated gene fusion at the DNA level and subsequent production of fusion mRNA. In these studies, successful fusion gene formation following transfection was dependent on concentrations of testosterone, estrogen, and progesterone, hormones known to alter chromosomal looping. Gupta et al. postulated that fusion RNAs may invade chromosomal DNA and bind the complementary sequences of two distant parent genes if the genes are in close physical proximity, such as through these hormone-induced changes in chromosomal looping. The resulting RNA/DNA hybrids would then be resolved through DNA repair mechanisms, resulting in recombination events that produce the final gene fusion at the chromosomal level.[18,19] However, the validation and generalizability of this mechanism requires more extensive research.

The types of genes involved in gene fusions, and the regions where breakpoints tend to occur, share some general features that may inform the mechanisms driving these mutations. The breakpoints of recurrent fusion genes in cancer tend to occur in specific intronic regions, rather than distributing randomly throughout the entire gene.[20] This non-random distribution of breakpoints appears to be influenced by the spatial proximity of chromosomes in the nucleus and features of the DNA sequence itself, such as repeats, fragile sites, and endonuclease misrecognition sites.[20] Breakpoints of oncogenic fusion genes tend to arise in intrinsically disordered regions of the genome, such as sites of transposable elements that function as “hot spots” for homologous recombination, other recombination-promoting sites like those associated with V(D)J recombination, sites of translin and topoisomerase II activity that play roles in DNA unwinding, and

sites with altered chromatin organization.[21] Fusion breakpoints also tend to preserve the reading frames of the flanking exons to maintain in-frame translation.[3,20] In a study of 583 gene fusions, the reading frames on the 5' and 3' sides of the breakpoint were compatible in 93%, even when the breakpoint occurred in the middle of an exon.[3] The significance of the 5'-3' arrangement of genes in a fusion remains a matter of debate. One study of 770 fusion genes of hematopoietic malignancies found that fusion genes tended to be overexpressed due to a combination of increased promoter activity supplied by the 5' partner gene and increased stability of the 3' partner gene caused by loss of its 3' UTR regulatory elements.[22] Another study analyzing six publicly available databases of cancer-associated gene fusions found that 5' partners tended to be more ubiquitously expressed and were more likely to be tumor suppressor genes while many 3' partners were more tissue-specific and tended to be proto-oncogenes.[23] However, a more recent study comparing the molecular signatures of partner genes across approximately 2200 gene fusions found that genes involved in fusions tended to be more highly and ubiquitously expressed, have higher protein-protein interaction network centrality and more molecular functions, possess more exons and isoforms, and contain many disorder-related elements like post-translational modifications, but a computational model could classify 3' and 5' parent genes by their gene functions and protein features at a rate only slightly higher than chance.[20] This suggests that fusion events may simply incorporate two "parent-like" genes, with the specific 5' or 3' ordering of those genes being less relevant.

Methods for fusion gene detection

Given their importance in cancer, several methods have been developed over the last few decades to allow for more efficient identification and characterization of fusion genes. Traditional methods were largely limited to guided approaches like chromosome banding, where Giemsa staining is used to visualize chromosome regions based on their unique band patterns, and fluorescence in-situ hybridization (FISH), in which probes for chromosome regions are hybridized to

fluorophores to allow visualization of structural rearrangements.[1,2] Limitations of these methods include the requirement of cultured cells and, in the case of chromosome banding, the inability to identify fusions arising in the same chromosome band.[1] These DNA-based methods also cannot detect fusions arising at the RNA level from cis- or trans-splicing.[2]

More recently, next-generation deep sequencing techniques opened the door for unbiased fusion gene discovery. Deep sequencing offers several key advantages over traditional chromosome banding. It can detect fusions arising from a variety of mechanisms, including intra-chromosomal rearrangements within the same chromosome band and, in the case of RNA-sequencing (RNA-seq), fusions arising from alternative splicing events. Deep sequencing also offers a more unbiased approach to fusion gene discovery. Before next-generation sequencing, the idea that fusion events most often involved a few key "promiscuous" genes was more prevalent, largely due to the use of FISH probes or one-sided polymerase chain reaction (PCR) methods that biased analyses toward the discovery of alternative partners for known fusion partner genes.[1] Deep sequencing challenged this paradigm by demonstrating that the vast majority of fusions occur between two genes that are not linked to any other fusion partner genes.[1] Fusion discovery by next-generation sequencing has its own disadvantages, however. Its accuracy is highly dependent on error-free library preparation, high quality reads, and accurate reference genome assemblies.[1] Given the tendency for error, validation of computationally predicted fusion genes by reverse transcription (RT)-PCR is standard protocol.[2] Additionally, the high sensitivity of computational fusion gene prediction often results in the identification of hundreds to thousands of fusion gene candidates per sample, many of which represent passenger mutations or readthrough artifacts, making differentiation of likely driver fusion genes a challenging and laborious process. Generally, fusion gene candidates that recur across multiple tumor samples are considered to have the most potential to be driver mutations, although other factors, such as the quality of the reads supporting the fusion and structural features of the

fusion, are also important considerations. Screening normal tissues for fusion gene candidates can be another effective measure for distinguishing oncogenic fusion candidates from those unlikely to be involved in malignancy.

Computational fusion gene discovery is often performed using RNA-seq data to identify candidate fusion transcripts. With paired-end reads, support for a fusion transcript generally comes in two forms: *spanning reads* and *split reads*. Spanning reads refer to the alignment of one paired-end read to gene A of a fusion transcript, while the other paired read aligns to gene B of a fusion transcript (Figure 46).[24] A split read refers to a read that is partially aligned to gene A and partially to gene B of a fusion transcript; these reads are useful pieces of evidence for pinpointing the exact fusion breakpoint (Figure 46).[24] The length of a split read's partial alignment to each gene in a fusion is the *anchor length*, and typically a minimum length of 10 base pairs is enforced by fusion callers to minimize false positive calls (Figure 46).[24]

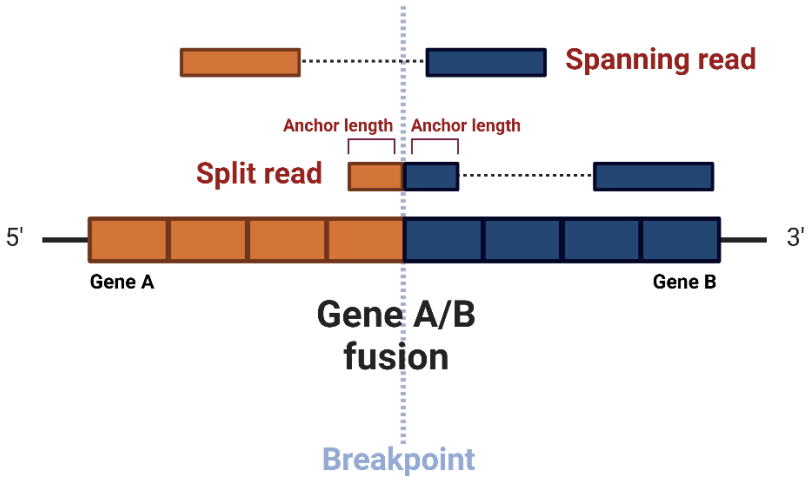


Figure 46: Fusion gene evidence from RNA-seq reads. Supporting reads for a fusion transcript are composed of spanning reads, in which one read aligns to gene A and the other paired read aligns to a distant gene B, and split reads, in which a read is partially aligned to gene A and partially aligned to gene B. For split reads, the length of the partial alignment to each gene is referred to as the anchor length, and this is generally required to be at least 10 base pairs long to meet criteria for a supporting read. Adapted from Liu et al. (2016). Created in BioRender. Owens, E. (2025) https://BioRender.com/l18n947.

While multiple algorithms have been developed that vary slightly in their approach to predicting fusion genes from RNA-seq data, the general principles employed by each tool are similar, as

reviewed in Wang et al.[25] Initially, reads are aligned to a reference genome or transcriptome, and read pairs that successfully align to the reference are excluded from further analysis while unmapped reads (which may represent split reads of a fusion) or discordantly mapped reads (which may represent spanning reads of a fusion) are kept for further evaluation. Discordant alignments get organized into groups based on shared pairs of breakpoints, and these groups are used to estimate fusion boundaries. Unmapped reads then get split into multiple segments, with the first and last segments mapped against the reference independently. If the two end segments map successfully to two different chromosomes or genes, the precise location of the fusion junction is then identified by adjusting the fragment boundaries and realigning to the reference (Figure 47). Fusion callers employ multiple filters to exclude reads unlikely to be involved in fusion events, such as those harboring ribosomal or mitochondrial RNA, or regions of repetitive sequences or sequence similarity. Final fusion gene candidates are also scored according to mapping quality, number of supporting reads, and read depth to minimize false positive calls.

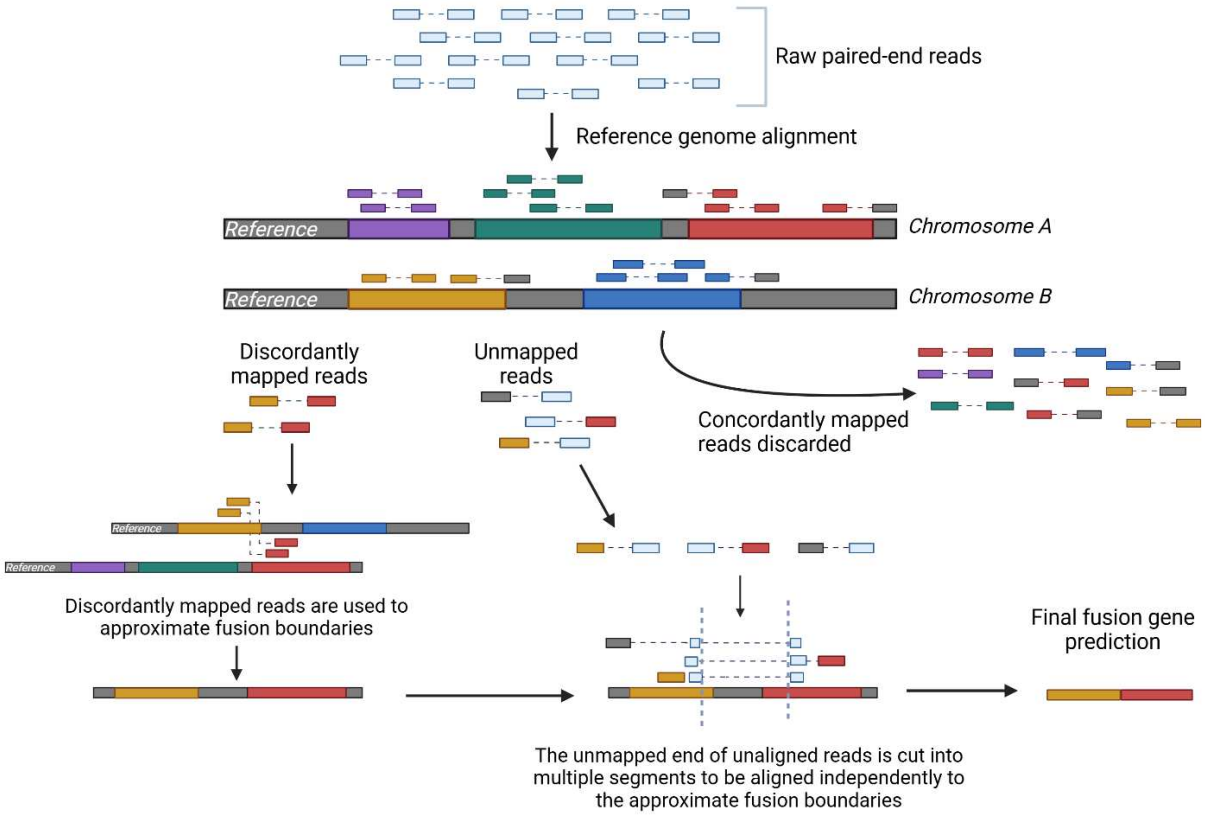


Figure 47: General approach for predicting fusion genes from RNA-seq reads. While individual fusion calling algorithms vary somewhat in their approach, they follow many of the same general principles. Generally, raw reads will be mapped to a reference genome, and those that remain unaligned or discordantly mapped will be kept for further analysis to identify fusion genes. Discordantly mapped reads, in which each read in a pair is able to be mapped to the reference but in unexpectedly distant locations, are used to approximate fusion boundaries. For unmapped reads, the unmapped end(s) is/are separated into smaller segments and independently re-aligned to the reference. If the smaller segments map to distant regions of the reference genome, then the fragment boundaries are adjusted and re-alignment performed again until the precise fusion breakpoint can be identified. Adapted from Wang et al. (2013). Created in BioRender. Owens, E. (2025) https://BioRender.com/r61z857.

Despite being a common technique for predicting fusion genes, the utilization of RNA-seq data has its limitations. Lowly expressed gene fusions or those that result in transcriptional silencing may be difficult or impossible to identify.[1] As mentioned previously, fusion RNAs presumably arising from alternative splicing events without a concurrent chromosomal aberration are frequently detected in non-cancerous tissues, while chromosome rearrangements are an attribute of cancer cells not expected to be present in normal cells. Thus, rigorous fusion gene discovery should involve evaluation of DNA evidence as well as RNA evidence. One recently emerging technique for detecting large structural variants like fusion genes in DNA without requiring the use of cultured cells is *long-read sequencing*. In contrast to the short-read sequencing performed by Illumina

sequencing platforms, which generate 150-bp read lengths, long-read sequencing methods generate reads several thousand base pairs long,[26] enabling a more complete picture of the structural organization of the genome and the detection of any genomic structural variants.

Long-read sequencing via the single-molecule, real-time (SMRT) approach uses DNA polymerase to directly observe the synthesis of a DNA molecule in real time (Figure 48A).[26–28] In this process, hairpin loop adapters are attached to each end of a double stranded DNA molecule and loaded onto a SMRT cell. Scattered throughout the SMRT cell are millions of wells tens of nanometers in diameter termed *zero-mode waveguides* (ZMWs) that each contain a single DNA polymerase molecule anchored to the bottom by a biotin/streptavidin interaction. These ZMWs are illuminated from below, and their small size causes the 600 nm-wavelength visible light to exponentially decay, illuminating only the bottom 30 nm of the ZMW to achieve the signal-to-noise ratio required for accurate base calling from a single 10 nm-wide DNA polymerase molecule. One of the adapters on a single molecule of DNA binds the polymerase to immobilize the template DNA in the ZMW for sequencing. Fluorophore-labeled nucleotides then flood the cell and diffuse down into the ZMWs. When the correct nucleotide is detected by the DNA polymerase, it is incorporated into the growing strand and emits a colored light signal that is recorded by the instrument as the base call for that position. Because the hairpin loop adapters create a closed circle of template DNA, the polymerase can continue on to the opposite strand after replicating the first strand of template DNA, sequencing both strands in a rolling circle fashion as a single continuous long read (Figure 48B). Multiple passes over both strands of the template DNA are performed to minimize carryover of any sequencing errors generated in a single pass. A standard high fidelity (HiFi) read undergoes 10 passes, which corresponds to a 99.9% consensus accuracy.[29] These HiFi reads can then be aligned to a reference genome to identify large structural variants.

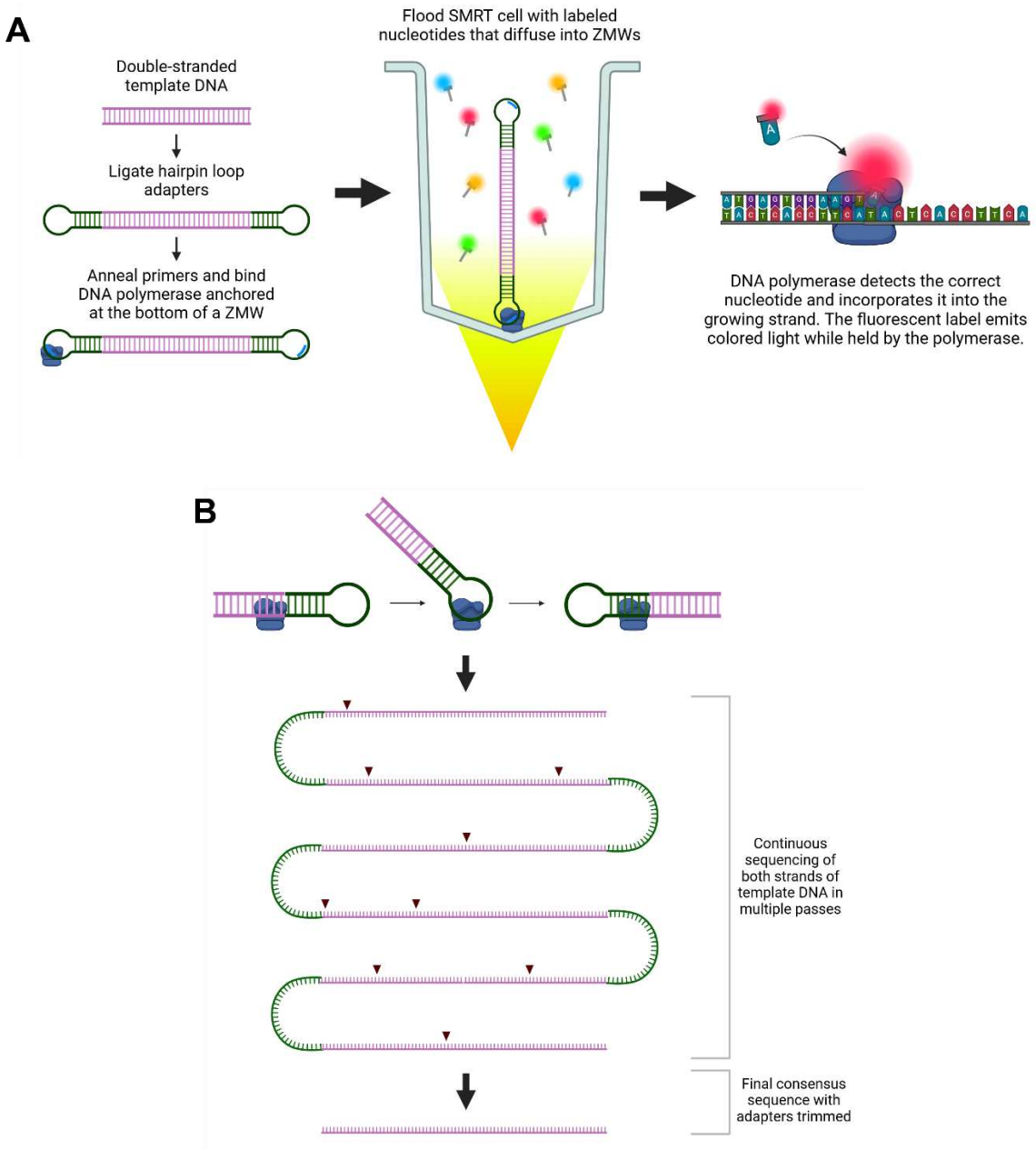


Figure 48: An overview of single-molecule, real-time and circular consensus sequencing. (**A**) SMRT sequencing. A single molecule of template DNA is sequenced in real-time using fluorophore-labeled nucleotides and a single DNA polymerase molecule anchored in a zero-mode waveguide (ZMW) illuminated from the bottom. Created in BioRender. Owens, E. (2025) https://BioRender.com/o00i653. (**B**) Consensus circular sequencing (CCS). The hairpin loop adapters allow both strands of the template DNA to be sequenced in a continuous circle, producing long continuous reads in multiple passes. A final consensus sequence is made after 10 passes to resolve errors that may arise in any single sequencing pass (indicated by the red arrows). Created in BioRender. Owens, E. (2025) https://BioRender.com/f51h647.

Fusion genes and other structural variants in canine PTCL

Currently, our understanding of fusion genes in canine PTCL is limited to unpublished data performed by a previous member of our lab, Dr. Lauren Harris. She utilized the fusion calling

algorithm FusionCatcher (version 1.10) to predict 2,399 tumor-specific fusion gene candidates in bulk RNA-seq data from the 26 cases of $CD4^+$ PTCL. After filtering to include only fusion candidates involving exonic regions, coding sequences (CDS), and/or untranslated regions (UTR), candidates that were recurrent (i.e., the 3' fusion partner was identified in 2 or more $CD4^+$ PTCLs), or candidates that involved at least one cancer-associated gene according to the COSMIC cancer-gene census,[30] she identified 43 recurrent 3' gene partners in these dogs. She then analyzed the structure of these fusions to predict which candidates were most likely to result in the translation of chimeric proteins or juxtaposition of one partner gene to regulatory elements of the UTR of the other partner gene. From her preliminary data, we were able to prioritize 23 fusion gene candidates for further validation and screening in a larger cohort of canine $CD4^+$ PTCL.

Additionally, to our knowledge, there has been no investigation into large structural variants at the DNA level in canine PTCL, which severely limits our understanding of how these mutations may be contributing to the pathogenesis of these tumors. Long-read sequencing to identify structural variants would not only offer insight into the mechanism behind the formation of fusion genes detected by RNA-seq in $CD4^+$ PTCL, but also would enable discovery of other variants, like large deletions, that may also be implicated in the pathogenesis of $CD4^+$ PTCL.

To address these knowledge gaps in our understanding of the impact of structural variants in canine $CD4^+$ PTCL, our study sought to: 1) validate previously predicted fusion gene candidates by PCR and Sanger sequencing, 2) identify additional fusion gene candidates in a larger cohort of $CD4^+$ PTCLs, 3) screen 120 cases of $CD4^+$ PTCL to determine how recurrent these predicted fusions are, and 4) perform long-read sequencing on a subset of $CD4^+$ PTCL to discover and characterize structural variants and determine if fusion transcripts have a corresponding detectable chromosomal aberration.

Materials and Methods

Sample collection

Fine needle tissue aspirates submitted to the Clinical Hematopathology Lab at Colorado State University for routine diagnostics and identified as PTCL by flow cytometry using the methods and antibody panels described in Seelig et al.[31] were considered for this study. The development and performance of these antibodies in dogs has been previously described.[32,33] Cases were diagnosed as $CD4^+$ PTCL by flow cytometry if they exhibited a homogeneous expansion of T cells with a $CD45^+CD3^+CD4^+CD5^{+/-}CD21^-CD34^-$ immunophenotype. A total of 120 samples diagnosed with these criteria were selected for fusion gene screening by reverse transcription polymerase chain reaction (RT-PCR). 96 of these samples were selected for bulk RNA-seq analysis and fusion gene prediction. 6 of these samples were selected for long-read sequencing. The site sampled was peripheral lymph node in 113 cases, mediastinal lymph node in 1 case, mesenteric lymph node in 2 cases, mediastinum in 3 cases, and peripheral blood in 1 case (Supplemental Table 16).

Control samples for bulk RNA-seq included sorted $CD4^+$ nodal lymphocytes and sorted $CD4^+$ thymocytes collected from the lymph nodes of 5 healthy female mixed breed dogs ranging from 7 months to 1 year in age designated for IACUC-approved medical device research (Supplemental Table 17). These dogs were confirmed to have no evidence of lymphadenopathy, lymphoma, or other morbidities by physical examination and gross necropsy. Lymph node cells from all 5 dogs and thymocytes from 2 dogs were sorted into $CD4^+$ T-cell and $CD4^+$ SP thymocyte subsets, respectively, by a Sony SH800 cell sorter (Sony Biotechnology). Cells were sorted using positive selection with anti-CD5 (clone YKIX322.3) and anti-CD4 (clone YKIX302.9) and were only utilized if the desired population comprised over 90% of the sorted sample after purity analysis. These samples were also utilized for fusion gene screening by PCR.

Control samples for fusion gene screening by PCR included the samples of sorted $CD4^+$ thymocytes and $CD4^+$ lymphocytes discussed above, and samples of whole blood submitted for hematologic evaluation to the Clinical Pathology department at Colorado State University from an additional 10 dogs. These dogs consisted of a Boxer, 2 Labrador Retrievers, a German Shorthair Pointer, and 4 mixed breed dogs and ranged from 2 to 12 years in age. 4 of these dogs were male, and 6 were female. These dogs did not have a reported history of lymphoma, although they did frequently have other morbidities. 2 dogs were diagnosed with immune mediated hemolytic anemia (IMHA), 1 was diagnosed with immune mediated thrombocytopenia (IMTP), and 1 had a diagnosis of mast cell tumor and nasal carcinoma. No morbidities were reported for the remaining 2 dogs. For 3 of these dogs, flow cytometry was performed on whole blood and no evidence of hematopathology was identified (Supplemental Table 17).

Fusion gene prediction from bulk RNA-seq data

For bulk RNA-sequencing, RNA was extracted, purified, and sequenced as previously described for Cohort 2 in the Materials & Methods section of Chapter One. Bulk RNA-seq reads were preprocessed, aligned to the CanFam3.1, ROS Cfam 1.0, CanFam4, and CanFam6 reference genomes, tabulated, normalized, and analyzed for differential gene expression as previously described in the Materials & Methods section of Chapter One. Dot plots of DESeq2 normalized count data were generated using GraphPad Prism for Windows, GraphPad Software, San Diego, California USA, www.graphpad.com (version 10.0.2) and means of normalized counts between experiment groups were compared by a Tukey's multiple comparisons test in GraphPad Prism. GSEA was performed on bulk RNA-seq data with clusterProfiler[38] (version 4.12.6) using the collective gene sets available through the Molecular Signatures Database (MSigDB).[39] The Wald test statistic was selected as the ranking metric for the list of differentially expressed genes in canine $CD4^+$ PTCL (as obtained by DESeq2) for its consideration of p-value in conjunction with log2 fold

change. clusterProfiler GSEA parameters were set to perform 10,000 permutations and enforce a p-value cutoff of 0.05 following correction using the Benjamini-Hochberg procedure.

Fusion calling was performed using STAR-Fusion (version 1.12.0) with default parameters and genome resource libraries built for CanFam3.1, ROS Cfam 1.0, CanFam4, and CanFam6 using the ctat-genome-lib-builder from the Trinity Cancer Transcriptome Analysis Toolkit (https://github.com/TrinityCTAT/Trinity_CTAT/wiki). Fusion pairs found in the control lymphocyte and thymocyte samples were discarded to identify tumor-specific candidates. Supporting read alignments were visualized using FusionInspector (version 2.10.0).[34] Fusion gene structure was evaluated using Geneious Prime 2023.2.1 (http://www.geneious.com/). Prediction of fusion gene impact on translation was performed using FusionCatcher, FusionInspector, and Geneious Prime.

Polymerase chain reaction

RNA was extracted from 120 $CD4^+$ PTCL samples and 17 samples of control dogs using the Purelink RNA Mini Kit (Thermo Fisher Scientific, Waltham, MA) and quality measured with a NanoDrop One UV-Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, MA). The mean A260/280 was 2.02, with a range of 1.06–2.2. The median A260/230 was 1.90, with a range of 0.02-2.42. First-strand cDNA synthesis was performed using the ProtoScript® First Strand cDNA Synthesis kit (New England Biolabs, Ipswich, MA). Successful cDNA synthesis was confirmed by PCR amplification of the wild-type *GNAS* gene (forward primer sequence: GATCCTGCATGTTAATGGGTTT; reverse primer sequence: ATCGCCTCTTTCAGGTTGTT).

Fusion genes predicted from FusionCatcher or STAR-Fusion were reconstructed from the CanFam3.1 and ROS Cfam 1.0 reference genomes using Geneious Prime 2023.2.1 (http://www.geneious.com/). The Geneious primer design tool was used to design PCR primers.

For any given fusion gene, the forward primer was constructed to bind approximately 20 bp of the 5' partner gene adjacent to the predicted fusion breakpoint, and the reverse primer was designed to bind approximately 20 bp of the 3' partner gene adjacent to the predicted fusion breakpoint. Primers had a mean GC content of 53.7% (range: 40.9-64%). Two versions of each reverse primer were constructed: one with a 6-FAM fluorescent dye for PCR reactions preceding capillary electrophoresis (CE), and one undyed version for PCR reactions preceding cloning (Table 9).

Table 9: PCR amplification primer sequences for $CD4^+$ PTCL fusion genes.

Fusion Gene	**Fusion Caller**	**5' primer sequence**	**3' primer sequence**
BZW1-HSPD1	STAR-Fusion	ACACCTCTTCCCTCCTCCTC	6FAM-GGACTTCCCCAACTCTGCTC
CHD3-EIF5A	STAR-Fusion	AGGACTACGAGGAGGACGAC	6FAM-TCCACGATCTTACATGGCCG
PER1-EIF5A	STAR-Fusion	ATTATGCAACCCGCCTCCTG	6FAM-TCCACGATCTTACATGGCCG
GRB10-IKZF1	STAR-Fusion	CGAAGAAGGAACCCATGGCT	6FAM-CTCCGCACATTCTTCCCCAT
GRB10-IKZFI	FusionCatcher	GGCTGCCCAGATTCCTTCTT	6FAM-TAGTTGCAGAGGTGGCACTT
MYC-TRIB1	STAR-Fusion	CGCATTTTGGACGCTGGATC	6FAM-GTCCCCGTGGTGTTCAGAAT
PTMA-HMGB1	STAR-Fusion	GTGCCCCATCATGTCAGACG	6FAM-GTCCGCCTTAGCCATGTCTT
PTMA-LMO4	STAR-Fusion	GTGCCCCATCATGTCAGACG	6FAM-TGTGCCAGTAGCTGTCCATG
REV3L-FYN	STAR-Fusion	CAGTGAGGATAGTGACCGCC	AGCAGTGGATCATGAGCTCG
YWHAZ-KLF10	STAR-Fusion	TTCCCGTCACTCAACCCATC	6FAM-CAGTAGATGGCGCTGATGC
KLF10-YWHAZ	STAR-Fusion	GTCGAGTGTCTCCGAGCG	6FAM-TACGGACTTCATGCAGGCTG
B2M-GNAS	FusionCatcher	CTGGCCTTGCTCCTCATC	6FAM-ATCGCCTCTTTCAGGTTGTT
JPT1-GATD3A	FusionCatcher	CGATCCCAACAGCAGGAATAG	6FAM-CTTCCTTCACACAATGCTTGG
MROH1-GATD3A	FusionCatcher	AGACGCGAGAGGAGTACAT	6FAM-CTTCCTTCACACAATGCTTGG
TPD52L2-GATD3A	FusionCatcher	AGACTCAGGAAACCCTCTCTC	6FAM-CTTCCTTCACACAATGCTTGG
MCM5-PTMA	FusionCatcher	AGCCCTTTGCCACAGAA	6FAM-CCTCATCAGTCTTCTGCTTCTT
NCL-PTMA	FusionCatcher	GACGTCGCTTGGATTCTTCT	6FAM-CCTCATCAGTCTTCTGCTTCTT
RGS10-PTMA	FusionCatcher	CCCGGCTCAATGAGAAGATA	6FAM-CCTCATCAGTCTTCTGCTTCTT
NFKB2-TRIM8	FusionCatcher	GTGAAGGAGGACAGTGCATAC	6FAM-CAGGAAGAGCTTGGAGTTCAG
SRSF5-LMO4	FusionCatcher	TGGAGAGGCGTAGACGATTT	6FAM-AGATGATACACATTGCCTTGGG
TOX2-LMO4	FusionCatcher	TGGCGCACCTGGACTATTA	6FAM-AGATGATACACATTGCCTTGGG

BCL11B-MYH9	FusionCatcher	TACTGCGGCAAGGTGTTCAA	6FAM-CCTTTGCAATGCTCTGGCTG
DENND4B-MYH9	FusionCatcher	AGTCCACTGAAGCCCTGGAA	6FAM-ACTGTAGTGGTGACAGCTGT
PCGF3-IKZF1	FusionCatcher	TGGAGATTCAAGAAGGCGCC	6FAM-TAGTTGCAGAGGTGGCACTT
ITM2C-XPO1	FusionCatcher	CGCAGCCATGGTGAAGATCA	6FAM-TGGGCTCCTTCTCCATGGTA
SPOCK2-XPO1	FusionCatcher	CAGCTGACCGTGAGATGTGA	6FAM-CCCCATCCCAGCCACAAAAT
TFAM-MYH9	FusionCatcher	ACATCACAGGTAAAGCTGAAGACT	6FAM-TCAGGGCCATGTTCTTCTGG
VBP1-JUN	FusionCatcher	GGTTGTGGCTTGGGAGAAGT	6FAM-TCTACGCCAACCTCAGCAAC

gBlock™ custom gene fragments (Integrated DNA Technologies, Coralville, IA) were used as positive template controls for PCR primers. For the fusion candidates predicted by FusionCatcher (Table 9), a separate gBlock™ was designed for each fusion and contained the predicted fusion sequence targeted by our PCR primers, plus additional nucleotides in the region spanned by the primers that were not present in the predicted fusion. These additional nucleotides allowed the amplicon size of the positive control gBlocks™ to be distinguished from true fusion events. For fusions with shared 3' partner genes, gBlocks™ were multiplexed to include sequences from all predicted 5' partner genes and one sequence from the 3' partner gene that encompassed all predicted fusion breakpoints. For the group of fusions predicted by STAR-Fusion (Table 9), all predicted fusion sequences targeted by our primers were multiplexed into a single gBlock™, with variable amplicon sizes resulting depending on the primer pair used.

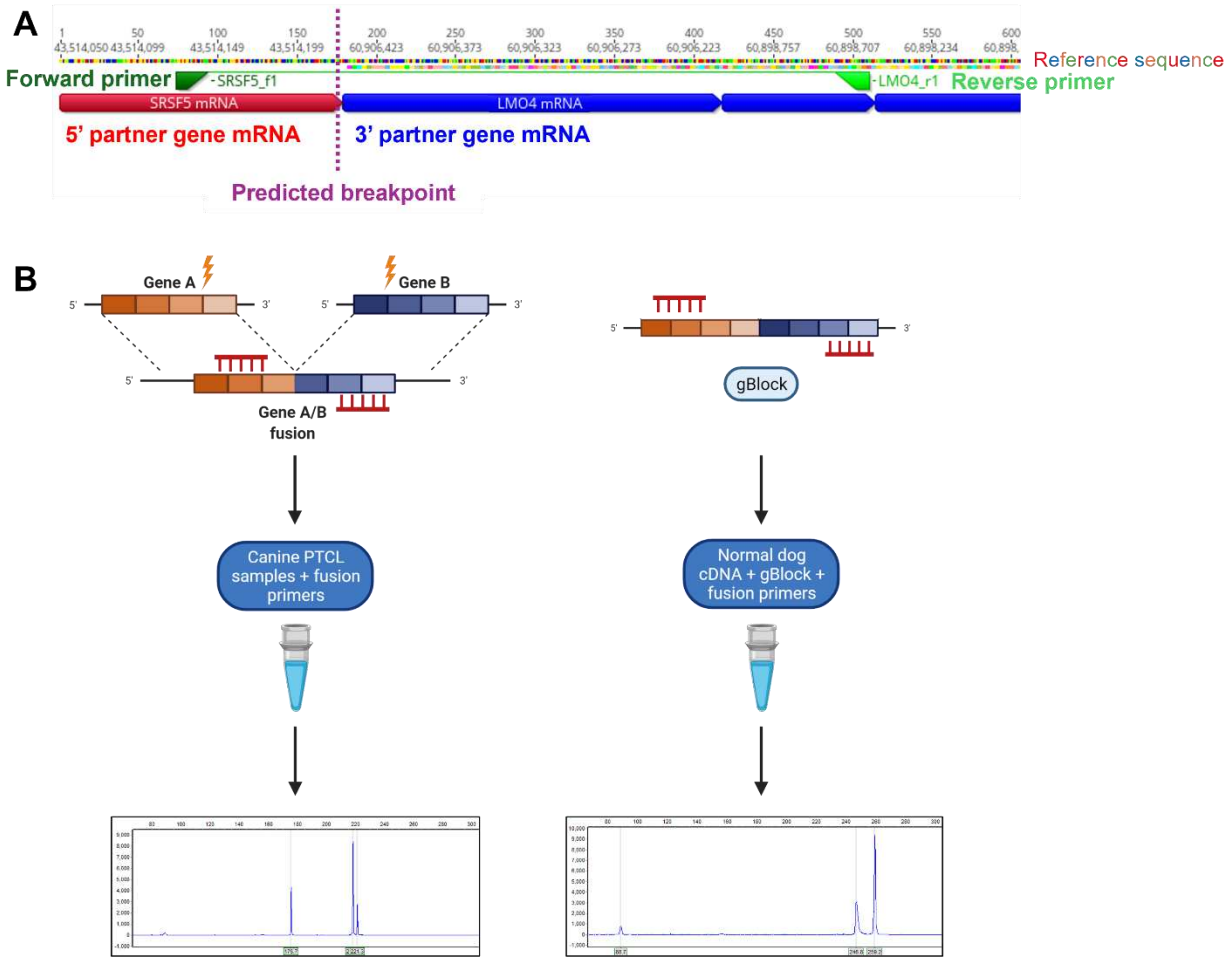


Figure 49: Overview of fusion gene primer and gBlock design. (**A**) Predicted fusions were reconstructed from reference genome sequences using Geneious Prime, and the Geneious primer design tool was used to identify candidate primers spanning the predicted fusion breakpoint (purple dotted line) by binding the partner gene sequences just 5' and 3' of the breakpoint. The *SRSF5-LMO4* fusion is shown here as an example. mRNA of the 5' partner gene *SRSF5* (shown in red) and mRNA of the 3' partner gene *LMO4* (shown in blue) are aligned to the reference genome (top). Above the reference sequence, the top numbers in gray represent the position within the fusion gene construct, while the bottom row of gray numbers denotes the original chromosome position of the parent genes. The forward primer sequence is annotated in dark green and the reverse primer sequence is annotated in light green above their complementary mRNA sequences, with the pointed ends indicating directionality. (**B**) PCR primers (red) were designed to target mRNA sequences adjacent to the breakpoints of the predicted gene fusions. These primers were then combined with canine $CD4^+$ PTCL or control dog template cDNA along with other PCR reaction reagents and amplified via RT-PCR. PCR products were analyzed by capillary electrophoresis. (**C**) gBlocks™ gene fragments containing the predicted fusion sequences as well as additional wild-type base pairs normally excluded from the fusion were used as positive template controls for PCR primers. These gBlocks™ were combined with fusion gene primer mixes and other PCR reaction reagents. The additional wild-type pairs result in an amplicon size that is distinguishable from that of the true fusion events.

PCR primer mixes were made using 2.5 uL of the forward primer, 2.5 uL of the reverse primer, and 45 uL TE. Master mixes for were made using 12.5 uL RNAse free water, 12.5 uL of 2x concentrated Multiplex PCR Master Mix (Qiagen, Hilden, Germany), and 1.25 uL of the appropriate primer mix. 25 uL of master mix was added to 1.5 uL of the appropriate template DNA (PTCL

cDNA, normal dog cDNA, or gBlock™ DNA fragments). PCR was performed on an Applied Biosystems MiniAmp™ Thermal Cycler (Thermo Fisher Scientific, Waltham, MA) with an initial denaturation step at 95° C for 15 min; followed by 40 cycles of denaturation at 94° C for 30 seconds, annealing at 56° C for 1 minute 30 seconds, and extension at 72° C for 1 minute 30 seconds; and concluded with a final extension step at 72° C for 20 minutes. Capillary electrophoresis of amplified PCR products was performed on an Applied Biosystems 3130xl Genetic Analyzer (Thermo Fisher Scientific, Waltham, MA), and GeneMarker version 1.97 (SoftGenetics, LLC., State College, PA) was used for visual fragment analysis.

Cloning and Sanger sequencing

For the *B2M-GNAS*, *NCL-PTMA, MROH1-GATD3A, SRSF5-LMO4, TOX2-LMO4, JPT1-GATD3A,* and *TPD52L2-GATD3A* gene fusions predicted by FusionCatcher that were detected in $CD4^+$ PTCL by our PCR assay, PCR products were re-amplified with undyed reverse primers and purified with the Clean & Concentrator-5 kit (Zymo Research, Irvine, CA). Purified PCR products were ligated into a PDrive cloning vector with the PCR cloning kit (Qiagen, Hilden, Germany) according to manufacturer instructions. Ligated DNA was added to a 100-µL tube of *Mix & Go!* TG1 Z-competent *E. coli* cells (Zymo Research, Irvine, CA) and incubated for 1 hour at 37° C and 250 rpm in Super optimal medium with catabolic repressor (SOC) medium before being spread onto solid media selective for ampicillin and kanamycin resistance and incubated overnight at 37° C. 1-8 successfully transformed colonies were identified by blue-white colony selection and incubated overnight at 37° C and 250 rpm in Lysogeny Broth (LB) broth with Ampicillin-100. Plasmid DNA was extracted according to manufacturer instructions with the Zyppy Plasmid Miniprep kit (Zymo Research, Irvine, CA) and combined with a 1:20 dilution of T7 primer (Thermo Fisher Scientific, Waltham, MA) in individual wells of a 96-well plate. Sequencing master mix was then prepared according to manufacturer instructions using the BigDye Terminator v3.1 Cycle Sequencing kit

(Thermo Fisher Scientific, Waltham, MA) and added to each well. Cycle sequencing was then performed on an Applied Biosystems MiniAmp™ Thermal Cycler (Thermo Fisher Scientific, Waltham, MA) using the following protocol: 95° C for 2 min 30 sec (x1), 95° C for 10 sec, 50° C for 8 sec, 60° C for 2 min 30 sec (x20), 95° C for 10 sec, 49° C for 8 sec, 60° C for 2 min, 20 sec (x10), 95° C for 10 sec, 52° C for 8 sec, 60° C for 2 min, 20 sec (x10), and held at 4° C. Sequencing reactions were purified with ethanol/EDTA precipitation and resuspended in Hi-Di™ Formamide (Thermo Fisher Scientific, Waltham, MA). Capillary electrophoresis was performed using an Applied Biosystems 3130x Genetic Analyzer, and sequencing electropherograms were analyzed using Geneious Prime 2023.2.1 (http://www.geneious.com/).

Long-read DNA sequencing

DNA was extracted from 6 $CD4^+$ PTCLs (Supplemental Table 18) using the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany). Sample concentrations and purity were measured with a NanoDRop One UV-Vis Spectrophotometer (Thermo Fisher Scientific, Waltham, MA), and DIN scores and amounts of high molecular weight DNA were measured with a 4150 TapeStation system using Genomic DNA ScreenTapes (Agilent Technologies, Inc., Santa Clara, CA). The median concentration of these samples was 300 ng/uL, with a range of 90-1530 ng/uL. The median A260/280 of these samples was 1.93, with a range of 1.91-1.97. The median A260/230 was 2.4, with a range of 2.1-2.4. The median DIN was 6.75, with a range of 6.6-7.9. The median percentage of DNA in each sample that exceeded 20 kb in size was 29%, with a range of 8-49%. Samples were submitted to the Arizona Genomics Institute for additional quality analysis, HiFi Single Molecule, Real-Time (SMRT) library construction, and HiFi long-read sequencing on a Revio platform (Pacific Biosciences of California, Inc., Menlo Park, CA).

Raw HiFi reads were aligned to the CanFam3.1 (accessed through Ensembl release 104[40]) and CanFam4 (UU_Cfam_GSD_1.0, referenced through Ensembl release 109[41]) canine reference

genomes with pbmm2 (version 1.14.99, https://github.com/PacificBiosciences/pbmm2), a customized SMRT wrapper for minimap2.[42] Structural variants were called using pbsv (version 2.9.0, https://github.com/PacificBiosciences/pbsv). Bed files of tandem repeat locations for the CanFam3.1 and CanFam4 reference genomes were acquired from UCSC Genome Browser[43] and provided to pbsv to improve variant calling in repetitive genomic regions. Variant calls were filtered and annotated with svpack (https://github.com/PacificBiosciences/svpack). "PASS" structural variants at least 50 bp long were retained for downstream analysis. Resulting variant call files (VCF) were analyzed and visualized in R[36] (version 4.4.0) and plots for data visualization were generated using ggplot2[35] (version 3.5.1). JBrowse[44] (version 2.18.0) was used to draw circos plots and visualize read alignments and structural variants. The bioinformatics pipeline and software details for structural variant calling from HiFi reads have been made available in a public Github repository: https://github.com/pirateyoho/Canine-PTCL-Long-Read-Seq-Variant-Analysis.

Results

Multiple recurrent fusion genes were identified in $CD4^+$ PTCL

The fusion gene candidates from Dr. Lauren Harris' preliminary data of 26 $CD4^+$ PTCLs that we initially selected for additional study included *B2M-GNAS, BCL11B-MYH9, DENND4B-MYH9, GRB10-IKZF1, ITM2C-XPO1, JPT1-GATD3A, MCM5-PTMA, MROH1-GATD3A, NCL-PTMA, NFKB2-TRIM8, PCGF3-IKZF1, RGS10-PTMA, SPOCK2-XPO1, SRSF5-LMO4, TFAM-MYH9, TOX2-LMO4, TPD52L2-GATD3A,* and *VBP1-JUN* (Table 10, Supplemental Table 19). These fusion candidates were selected from the list of the most recurrent 3' fusion partner genes identified by FusionCatcher and tended to have either a strong number of supporting reads or structural features, such as the preservation of coding sequences, that supported a chance for formation of functional fusion protein.

STAR-Fusion identified 2,609 total tumor-specific fusion events in bulk RNA-seq data of 96 $CD4^+$ PTCL and 7 control samples of normal $CD4^+$ nodal lymphocytes and $CD4^+$ thymocytes. When reads were aligned to the CanFam3.1 reference genome, 1,099 total tumor-specific fusion events were identified across all $CD4^+$ PTCL samples, and these involved 289 unique fusion genes. When reads were aligned to the CanFam4 reference genome, 731 total tumor-specific fusion events involving 305 unique fusion genes were identified across all samples. When reads were aligned to the ROS Cfam 1.0 reference genome, 779 total tumor-specific fusion events involving 317 unique fusion genes were discovered across all samples.

Fusion genes that were called by STAR-Fusion in multiple samples and were consistently predicted regardless of reference genome used were prioritized for validation by PCR. These included *BZW1-HSPD1*, which was called in the same 11 dogs using all 3 reference genomes and 1 additional dog using the Canam 3.1 alignment; *PTMA-HMGB1*, which was called in the same 9 dogs using all 3 reference genomes and 1 additional dog using the CanFam4 alignment; *REV3L-FYN*, which was called in the same 4 dogs across all 3 reference genomes; *PER1-EIF5A*, which was called in the same 4 dogs using all 3 reference genomes; and *CHD3-EIF5A*, which was predicted in the same 3 dogs using both the ROS Cfam 1.0 and CanFam4 reference genomes. Though not recurrent like these candidates, the *MYC-TRIB1* fusion was also selected for validation by PCR for its implication of prominent oncogene *MYC,* despite only being predicted in 1 dog by STAR-Fusion. Finally, fusion genes that were predicted by both FusionCatcher and STAR-Fusion were also prioritized for validation by PCR. These included *NCL-PTMA, PTMA-LMO4, PTMA-HMGB1, KLF10-YWHAZ, PER1-EIF5A, GRB10-IKZF1,* and *YWHAZ-KLF10* (Table 10, Supplemental Table 19). Evidence for these fusions included a combination of split reads and spanning reads, with the exception of *MYC-TRIB1* and *BZW1-HSPD1*, which were supported only by split reads (Table 10, Figure 50).

Table 10: Fusion gene candidates selected for validation by PCR from FusionCatcher and STAR-Fusion outputs. Breakpoints are based on the CanFam3.1 reference genome, with the exception of *CHD3-EIF5A*, which is based on the ROS Cfam 1.0 reference genome. UTR = untranslated region. CDS = coding sequence.

Fusion Gene	Fusion Caller	Reference genomes identified in	5' Breakpoint	3' Breakpoint	Maximum supporting split reads	Maximum supporting spanning reads	Protein translation effect
REV3L-FYN	STAR-Fusion	CanFam3.1, ROS Cfam 1.0, CanFam4	12:67953124:-	12:68106497:-	12	1	In-frame
NCL-PTMA	Fusion Catcher & STAR-Fusion	CanFam3.1	25:43340008:-	25:43566003:+	9	20	In-frame
PTMA-LMO4	Fusion Catcher & STAR-Fusion	CanFam3.1	25:43563385:+	6:60906444:-	7	7	In-frame
SRSF5-LMO4	Fusion Catcher	CanFam3.1	8:43514227:+	6:60906444:-	3	8	UTR fusion to complete in-frame CDS
TPD52L2-GATD3A	Fusion Catcher	CanFam3.1	24:47363544:+	31:38035219:+	4	8	In-frame
MYC-TRIB1	STAR-Fusion	CanFam3.1, ROS Cfam 1.0, CanFam4	13:25201185:+	13:23343857:+	14	0	Out-of-frame
B2M-GNAS	Fusion Catcher	CanFam3.1	30:11331865:+	24:43643409:+	6	6	In-frame
PTMA-HMGB1	Fusion Catcher & STAR-Fusion	CanFam3.1, ROS Cfam 1.0, CanFam4	25:43563385:+	25:9554184:+	39	16	Truncated CDS fusion to UTR
JPT1-GATD3A	Fusion Catcher	CanFam3.1	9:5321847:+	31:38035219:+	3	6	In-frame
KLF10-YWHAZ	Fusion Catcher & STAR-Fusion	CanFam3.1, ROS Cfam 1.0	13:4223542:-	13:2760450:-	7	13	Truncated CDS fusion to UTR
MROH1-GATD3A	Fusion Catcher	CanFam3.1	13:37690507:+	31:38035219:+	3	8	In-frame
TOX2-LMO4	Fusion Catcher	CanFam3.1	24:31523443:+	6:60906444:-	3	5	In-frame
PER1-EIF5A	Fusion Catcher & STAR-Fusion	CanFam3.1, ROS Cfam 1.0, CanFam4	5:32968315:-	5:32270433:+	16	7	Intronic fusion to complete out-of-frame CDS

BZW1-HSPD1	STAR-Fusion	CanFam3.1, ROS Cfam 1.0, CanFam4	37:10005293:+	37:7020338:-	28	0	UTR to complete out-of-frame CDS
CHD3-EIF5A	STAR-Fusion	ROS Cfam 1.0, CanFam4	5:32949209:+	5:32476762:+	11	3	Out-of-frame
BCL11B-MYH9	Fusion Catcher	CanFam3.1	8:67622861:-	10:28122190:+	3	8	In-frame
DENND4B-MYH9	Fusion Catcher	CanFam3.1	7:43194960:+	10:28122813:+	2	5	Truncated CDS fusion to UTR
GRB10-IKZF1	Fusion Catcher	CanFam3.1	18:1970944:-	18:1676646:+	3	38	In-frame
GRB10-IKZF1	Fusion Catcher & STAR-Fusion	CanFam3.1, ROS Cfam 1.0	18:1941001:-	18:1676646:+	6	9	Out-of-frame
ITM2C-XPO1	Fusion Catcher	CanFam3.1	25:42830052:+	10:61585054:-	3	5	Truncated CDS fusion to UTR
MCM5-PTMA	Fusion Catcher	CanFam3.1	10:28722768:-	25:43567242:+	2	4	In-frame
NFKB2-TRIM8	Fusion Catcher	CanFam3.1	28:14910267:+	28:15136432:+	26	457	In-frame
PCGF3-IKZF1	Fusion Catcher	CanFam3.1	3:91676726:-	18:1732840:+	2	3	UTR fusion to truncated CDS
RGS10-PTMA	Fusion Catcher	CanFam3.1	28:29688064:-	25:43563253:+	4	5	Truncated CDS fusion to UTR
SPOCK2-XPO1	Fusion Catcher	CanFam3.1	4:22752884:-	10:61542482:-	3	4	Truncated CDS fusion to UTR
TFAM-MYH9	Fusion Catcher	CanFam3.1	4:10597916:+	10:28110256:+	4	6	In-frame
VBP1-JUN	Fusion Catcher	CanFam3.1	X:123216276:+	5:50968371:-	4	9	Truncated CDS fusion to exon with no known CDS
YWHAZ-KLF10	Fusion Catcher & STAR-Fusion	CanFam3.1, ROS Cfam 1.0	13:2763430:-	13:4220105:-	19	13	UTR to truncated CDS

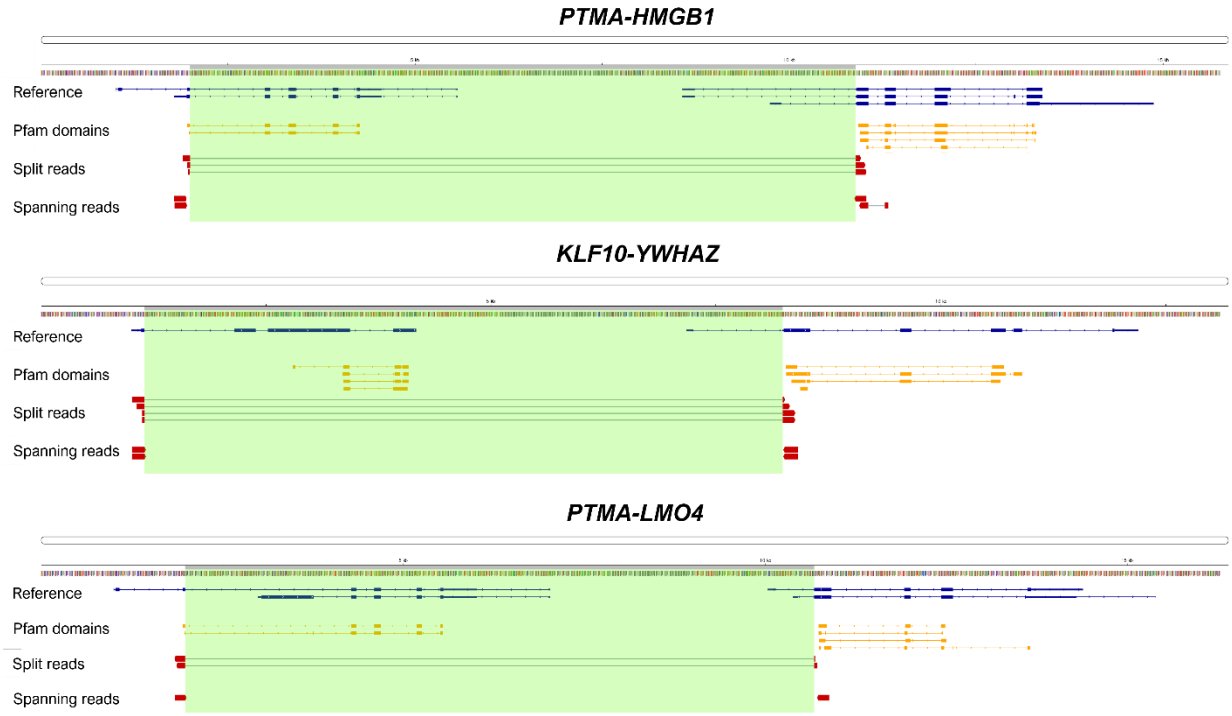


Figure 50: Visualization of supporting read alignments with FusionInspector. Examples using the *PTMA-HMGB1, KLF10-YWHAZ,* and *PTMA-LMO4* fusion genes show the alignment of RNA-seq reads (red) supporting the fusion as either split reads or spanning reads against the reference sequences for those genes (blue). Segments of the reference genes in the green highlighted region would be lost in the formation of the fusion gene. Protein coding domains as derived from Pfam database (http://pfam.xfam.org/) by FusionInspector are shown in yellow. Colored lines at the top are individual nucleotides of the reference sequence. See Figure 46 for illustration of split reads versus spanning reads.

Prediction of fusion impact with FusionCatcher, FusionInspector, and Geneious Prime revealed that the predicted *B2M-GNAS*, *BCL11B-MYH9*, *GRB10-IKZF1*, *JPT1-GATD3A*, *MROH1-GATD3A*, *NCL-PTMA*, *MCM5-PTMA, NFKB2-TRIM8, PTMA-LMO4*, *REV3L-FYN*, *TFAM-MYH9, TPD52L2-GATD3A*, and *TOX2-LMO4* fusions result in the creation of an in-frame sequence spanning both coding sequences of the partner genes, which may be translated into a novel fusion protein (Figure 51A-B). The *CHD3-EIF5A* and *MYC-TRIB1* fusions result in out-of-frame fusions that are unlikely to result in a functional fusion protein, although out-of-frame fusion RNAs may still function as regulatory long non-coding RNAs (lncRNAs).[2] For the *BZW1-HSPD1*, *DENND4B-MYH9, ITM2C-XPO1, KLF10-YWHAZ, PCGF3-IKZF1, PER1-EIF5A, PTMA-HMGB1*, *RGS10-PTMA, SPOCK2-XPO1, SRSF5-LMO4*, *VBP1-JUN,* and *YWHAZ-KLF10* fusions, the predicted breakpoint did not merge two coding regions, but instead involved fusion of UTRs or intronic regions to truncated or complete

coding sequences (Table 10). Fusions in this group that result in the truncation of coding sequences may interfere with normal protein translation. Fusions that fused intronic regions or UTRs to a complete CDS most often resulted in a frameshift that, while preserving the complete CDS, did not preserve the open reading frame (ORF). The exception was the *SRSF5-LMO4* fusion, which at the breakpoint predicted by FusionCatcher fuses the 3' UTR of *SRSF5* is to the complete CDS of the *LMO4* gene, with the ORF remaining intact (Figure 51B). This could place the expression of *LMO4* under the control of the regulatory elements of the 3' UTR of *SRSF5*.

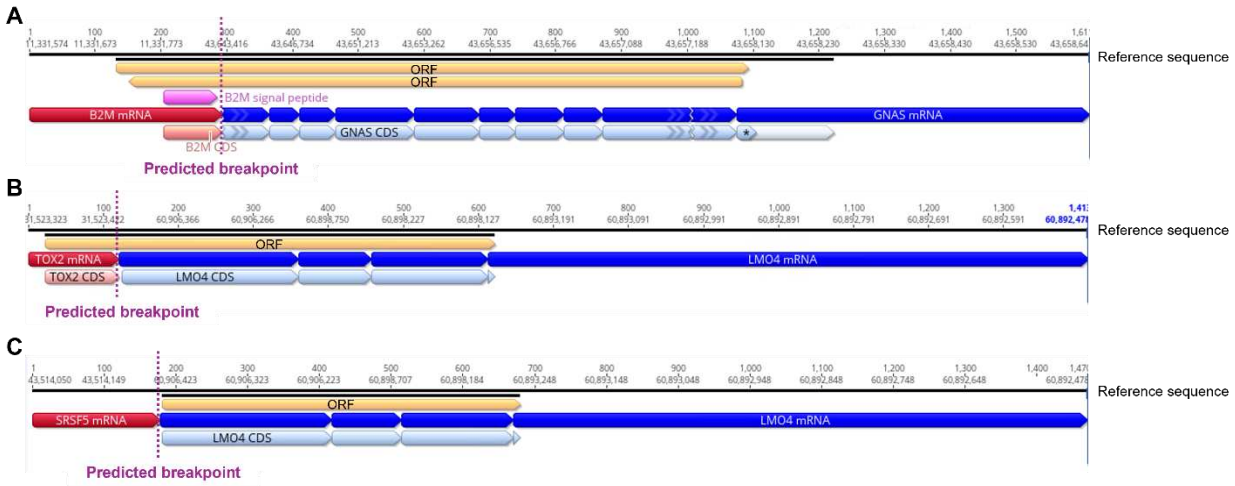
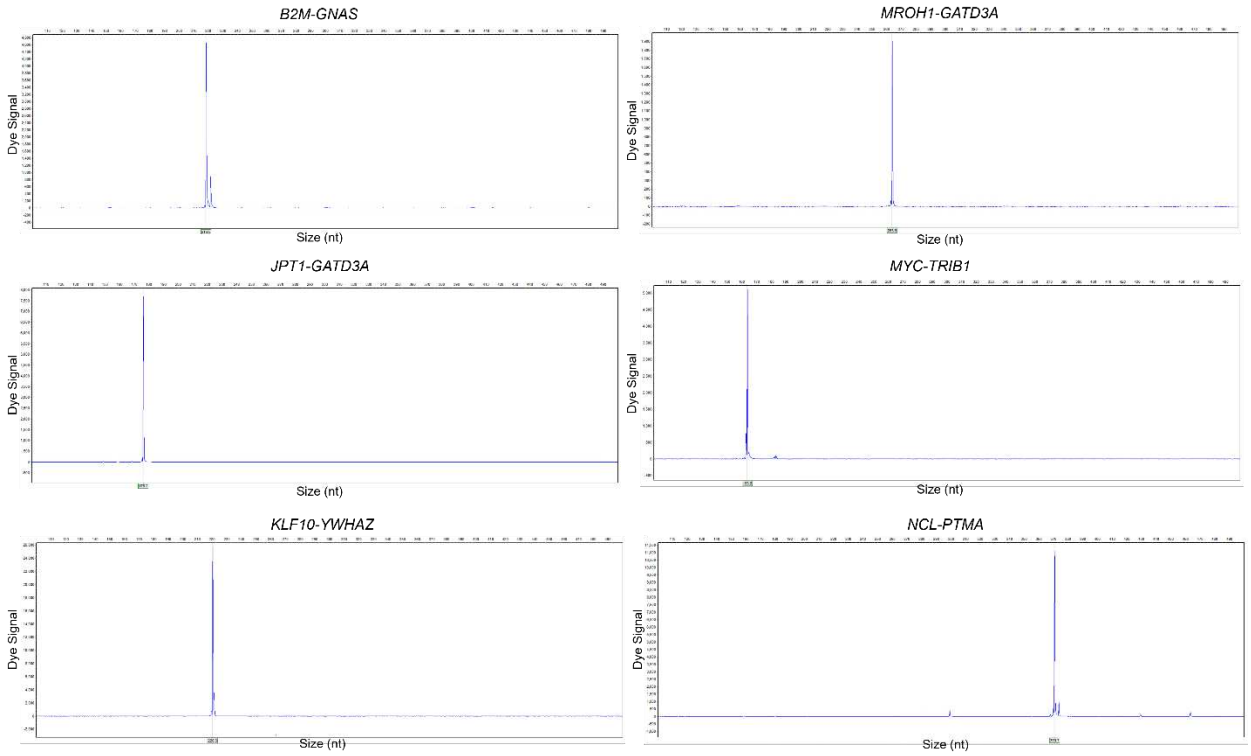


Figure 51: Predicted impacts of fusion genes on translation. The *B2M-GNAS* (**A**) and *TOX2-LMO4* (**B**) exemplify in-frame fusions that fuse the coding sequences of the parent genes together, which may result in the formation of novel fusion proteins. The *SRSF5-LMO4* fusion gene (**C**) fuses the 3' UTR of the *SRSF5* gene to the complete in-frame coding sequence of the *LMO4* gene, which could place expression of the *LMO4* gene under the control of regulatory elements in the UTR of *SRSF5*. mRNA sequences of 5' fusion partners are shown in dark red, with their associated coding sequences annotated in light red. mRNA sequences of 3' fusion partners are shown in dark blue, with their associated coding sequences annotated in light blue. Predicted breakpoints are denoted by the purple dotted line. Open reading frames are annotated above the mRNA sequences in yellow. The reference sequence is denoted by the solid black line above the mRNA sequences and ORF annotations. Above the reference sequence, the top numbers in gray represent the position within the fusion gene construct, while the bottom row of gray numbers denotes the original chromosome position of the parent genes.

Representative CE tracings for each fusion gene able to be amplified in $CD4^+$ PTCL samples by our fusion primers are shown in Figure 52.


B2M-GNAS
MROH1-GATD3A
JPT1-GATD3A
MYC-TRIB1
KLF10-YWHAZ
NCL-PTMA
Dye Signal
Size (nt)

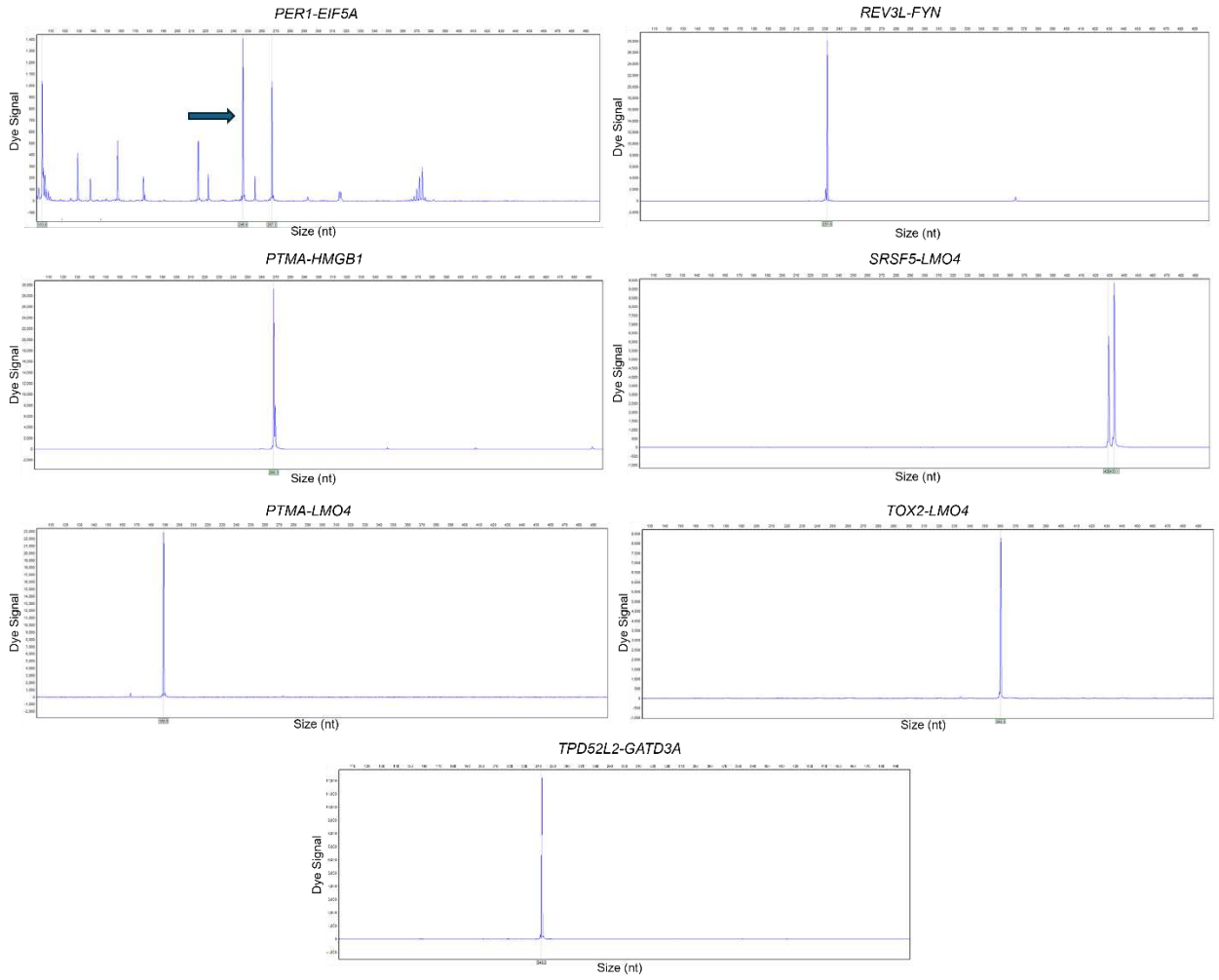


Figure 52: Representative CE tracings of PCR products amplified by each respective fusion gene primer. These results were consistent with the amplicon size predicted from breakpoint sequences derived from STAR-Fusion and FusionCatcher results.

For some fusion gene candidates, the predicted fusion amplicon was accompanied by additional PCR products of sizes that were consistent across numerous samples. For example, for the *TPD52L2-GATD3A* fusion gene, the predicted 241-bp amplicon size was accompanied by a 334-bp product in a subset of samples, a difference of 93 bp (Figure 53A). With cloning and Sanger sequencing, we confirmed that the 334-bp product was consistent with an addition exon from *GATD3A* being included, suggesting a splice variant of this fusion gene (Figure 53B-C).

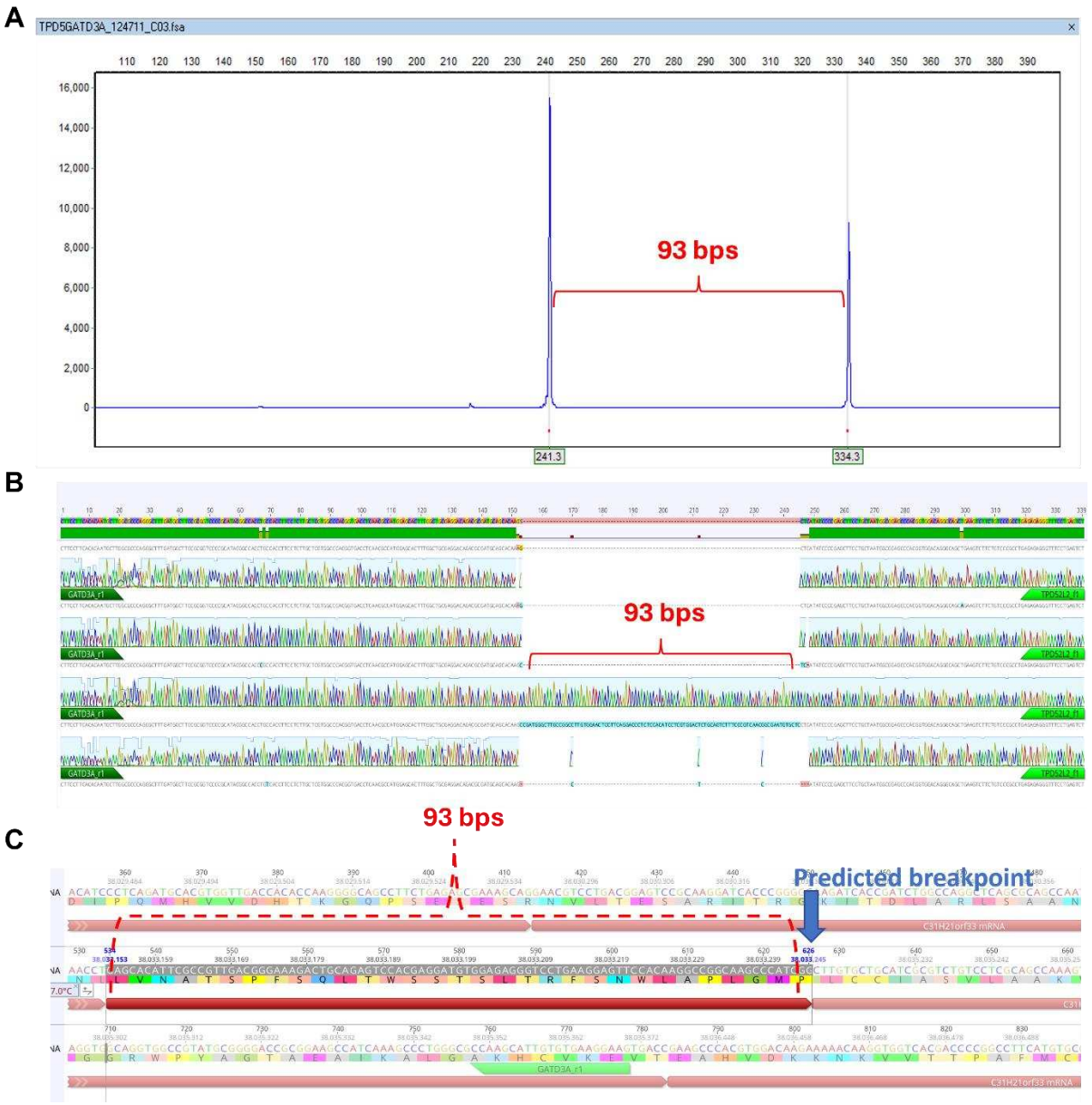


Figure 53: Splice variants of the *TPD52L2-GATD3A* fusion gene. (**A**) CE tracing of PCR products amplified with our *TPD52L2* forward primer and *GATD3A* reverse primer. The *TPD52L2-GATD3A* fusion gene was predicted to result in an amplicon of 241 bp with fusion primers designed to target the breakpoint sequence predicted by FusionCatcher. Some samples displayed an additional peak at 334 bp. (**B**) Sanger sequencing results from the same PTCL sample, visualized in Geneious Prime. Pictured at the top as a row of multicolored nucleotides is the reference sequence for the *TPD52L2-GATD3A* fusion gene as predicted by FusionCatcher. Below the reference sequence are the aligned chromatograms obtained through Sanger sequencing of 4 cloned PCR products from this sample. The corresponding forward primer and reverse primer sequences used for PCR amplification of fusion cDNA prior to cloning are annotated in dark green and light green below the chromatograms, respectively, with the pointed ends indicating directionality. The solid green line below the reference sequence indicates the consensus between the reference sequence and the aligned sequences from the cloned products. There is a 93-bp mismatch from the reference fusion sequence present in the third clone corresponding to the 334-bp PCR product appreciated in the CE tracing. (**C**) The 93 nucleotides present in the third clone mapped to the 93-nucleotide sequence of the *GATD3A (C31H21orf33)* exon 5' to the breakpoint predicted by FusionCatcher, suggesting that this exon was retained some transcripts of this fusion gene. Reference nucleotides and amino acid sequences are shown at the top. The red bars correspond to *GATD3A* mRNA exons. The *GATD3A* reverse

primer used for PCR amplification of this fusion gene candidate is annotated in light green below the reference sequence, with the pointed end indicating directionality.

While not all of the additional recurrent PCR products in our reactions were able to be validated by sequencing (likely due to chance colony selection since blue-white colony screening does not distinguish between ligated products), we suspect that most of these products represent splice variants like what was demonstrated in the case of *TPD52L2-GATD3A*. Still, only samples that displayed a peak at the originally predicted fusion amplicon size were counted as positive results for the purposes of our study.

Of the final list of 27 fusion gene candidates derived from our FusionCatcher and STAR-Fusion data, 13 (48%) were detectable in CD4$^+$ PTCL samples by PCR (Figure 54A, Supplemental Table 20). Of these 13 fusion genes detectable by PCR in CD4$^+$ PTCL, 10 were also detected by PCR in our control dog samples (Figure 54B, Supplemental Table 20). Tumor-specific fusion genes that were not detected in our control dog samples included *MROH1-GATD3A, TOX2-LMO4*, and *PER1-EIF5A*.

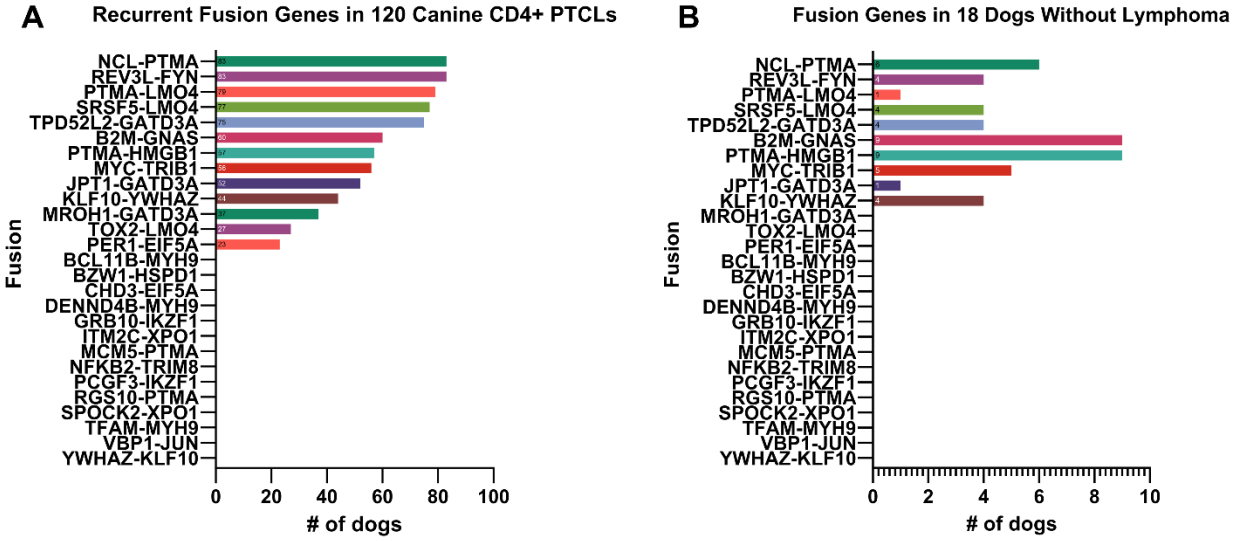


Figure 54: Summary of PCR results in CD4$^+$ PTCL (**A**) and control dogs (**B**) for fusion gene candidates predicted by FusionCatcher and STAR-Fusion.

While it varies by fusion gene, higher expression of fusion RNA in tumor samples than normal samples has been reported in humans, thought to be related to clonal expansion of fusion-expressing cells that normally make up only a low proportion of the cells in a tissue.[34] In our study, a significant difference in signal intensity of the amplified PCR product was only identified for the *REV3L-FYN* fusion gene, where signal intensity was significantly higher in tumor samples than samples from dogs without lymphoma (Figure 55). However, CE signal intensity is largely qualitative and not as sensitive as methods like quantitative PCR (qPCR) for measuring analyte concentrations, so future experiments utilizing qPCR should be considered to more accurately assess the differences in fusion transcript expression between CD4$^+$ PTCLs and normal samples.

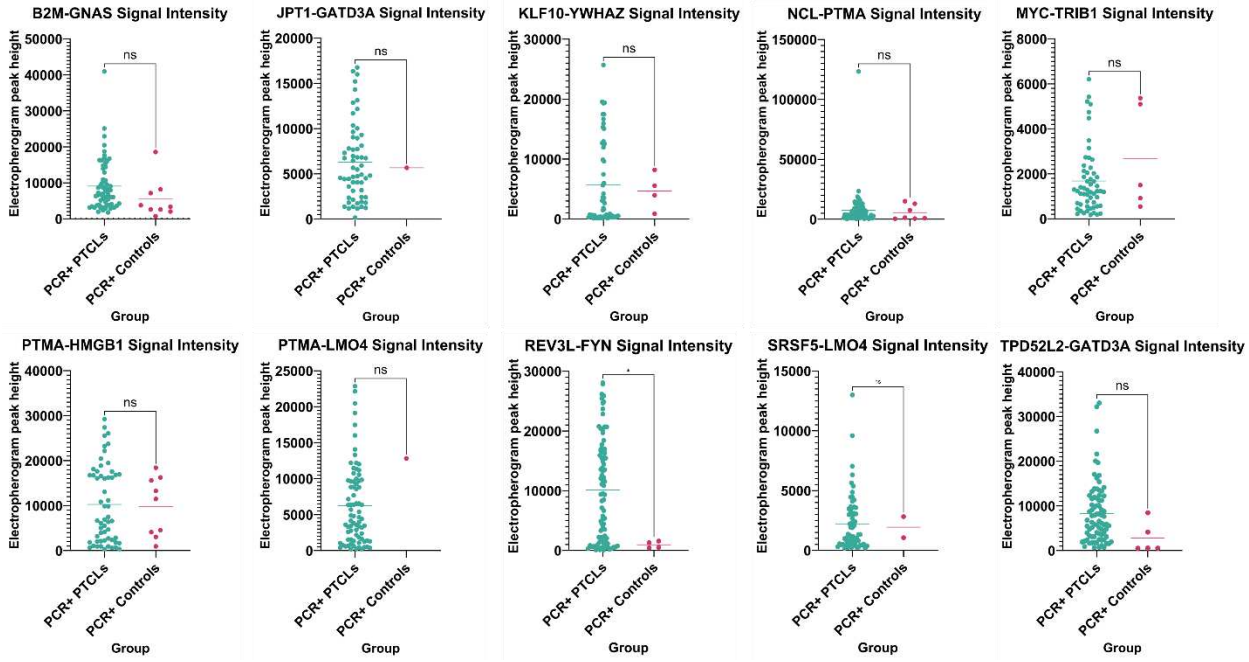


Figure 55: CE signal intensity for fusion genes detected in both CD4$^+$ PTCL and dogs without lymphoma. Signal intensity in CE experiments can function as a qualitative measurement of analyte concentration. No significant difference in CE signal intensity for a positive fusion detection by PCR was identified for the majority of fusion genes in our assay, with the exception of *REV3L-FYN*, where signal intensity was higher in CD4$^+$ PTCLs compared to control dogs without lymphoma, suggesting that *REV3L-FYN* fusion transcript expression may be higher in tumor samples.

Cloning and Sanger sequencing were performed as an additional validation step on a subset of the fusion gene candidates that were detectable by PCR, and the sequences of the amplified PCR

products were consistent with the fusion breakpoint sequences predicted by FusionCatcher (Figure 56).

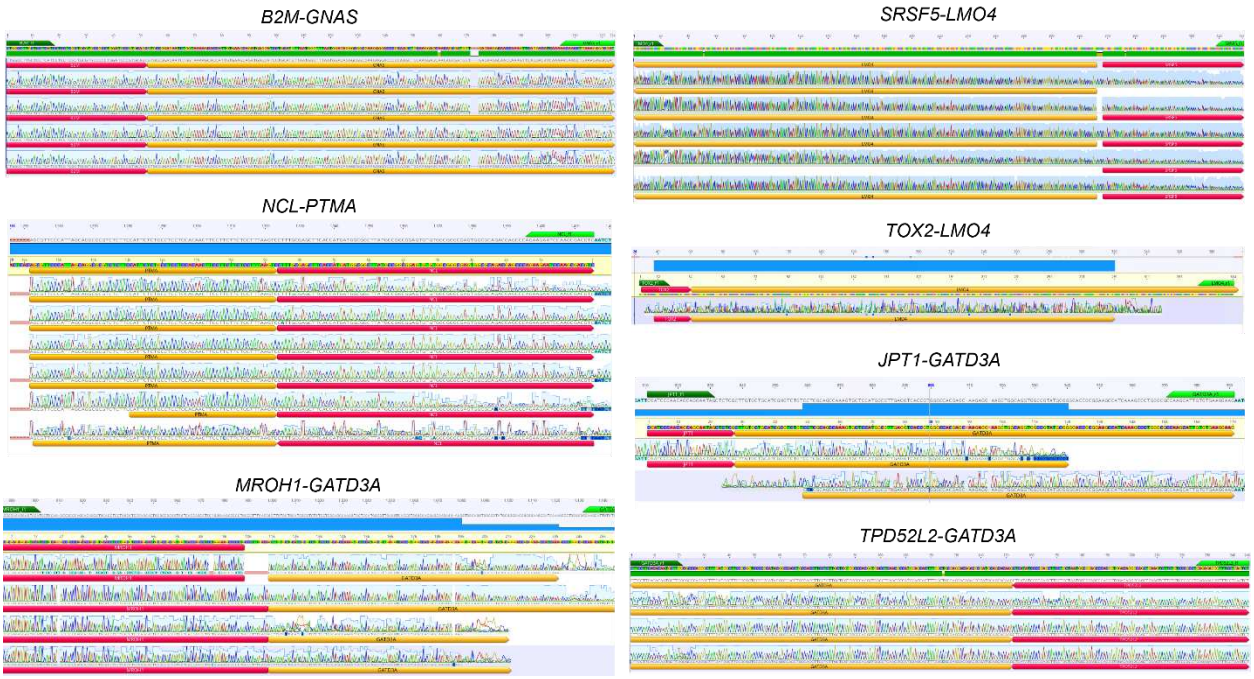


Figure 56: Cloning and Sanger sequencing of the *B2M-GNAS, NCL-PTMA, MROH1-GATD3A, SRSF5-LMO4, TOX2-LMO4, JPT1-GATD3A,* and *TPD52L2-GATD3A* PCR products validated the breakpoint sequences predicted by FusionCatcher. The reference sequence is shown as a line of multicolored nucleotides along the top of each figure. Below the reference sequence are the aligned chromatograms obtained through Sanger sequencing of extracted plasmid DNA. The corresponding forward primer and reverse primer sequences used for PCR amplification of fusion cDNA prior to cloning are annotated in dark green and light green above the reference sequence, respectively, with the pointed end indicating directionality. The solid dark green or blue line below the reference sequence indicates the consensus between the reference sequence and the aligned sequences from the cloned products. Sequences associated with the 3' fusion partner are annotated in red below the reference sequence and each corresponding chromatogram, and sequences associated with the 5' fusion partner are annotated in yellow below the reference sequence and each corresponding chromatogram.

The presence of fusion transcript is associated with altered partner gene expression in some $CD4^+$ PTCL fusions

One common mechanism by which fusion genes contribute to oncogenesis is through the dysregulated expression of one of the partner genes. For example, a fusion may bring one partner gene under the control of the regulatory elements of the other, as seen with cases of promoter exchange like the *STIL-TAL1* fusion in T-ALL. Alternatively, the fusion event may result in loss of the normal regulatory elements of one or both partner genes, as seen with the *CYCLIN D1-TROP2* fusion gene in a variety of human tumors, thereby dysregulating its expression. To investigate this

possibility in our CD4$^+$ PTCL fusion gene candidates, we performed bulk RNA-seq and fusion gene screening by PCR for 94 CD4$^+$ PTCL samples and 7 control samples of sorted CD4$^+$ lymphocytes and thymocytes and conducted differential expression analysis to compare the expression levels of each partner gene in PTCL samples where the fusion was present both to controls and to PTCL samples where the fusion gene was absent.

Regarding differential expression of 5' partner genes in our candidate fusion genes, the *B2M-GNAS, KLF10-YWHAZ*, *MROH1-GATD3A*, and *SRSF5-LMO4* fusions were associated with significantly higher expression of their respective 5' partner genes in CD4$^+$ PTCLs that expressed the fusion gene compared to those CD4$^+$ PTCLs that did not, based on both a Tukey's multiple comparisons test of normalized counts (Figure 57) and differential expression analysis performed by DESeq2 (Table 11). *MROH1-GATD3A* (discussed more extensively in the next section) was only detected in tumor samples in our study, while transcripts of *B2M-GNAS, KLF10-YWHAZ,* and *SRSF5-LMO4* were also identified in some control samples by PCR. Gene expression data was available for 1 control sample that was positive for *B2M-GNAS* fusion transcript by PCR; however, gene expression data was not available for the control samples that were PCR-positive for the *KLF10-YWHAZ* and *SRSF5-LMO4* fusions. *B2M* expression between a control sample PCR-positive for *B2M-GNAS* and CD4$^+$ PTCL samples expressing *B2M-GNAS* was not significantly different (Figure 57). *B2M* expression between this control sample and other control samples that were PCR-negative for the *B2M-GNAS* fusion was also not significantly different (Figure 57). Interestingly, CD4$^+$ PTCL samples that were PCR-negative for the *B2M-GNAS* fusion had significantly decreased expression of *B2M* compared to *B2M-GNAS*-negative control samples (Figure 57). A similar pattern was observed for *SRSF5-LMO4*: *SRSF5* expression was not significantly different between *SRSF5-LMO4*-expressing CD4$^+$ PTCLs and control samples negative for this fusion, but it was significantly increased compared to other CD4$^+$ PTCLs that lacked expression of this fusion mRNA by PCR (Figure 57).

These findings could suggest that *B2M* and *SRSF5* are normally downregulated in CD4$^+$ PTCL, but the presence of the *B2M-GNAS* fusion or *SRSF5-LMO4* fusion, respectively, stabilizes their expression back to normal levels. *B2M* is a component of the MHC class I molecule,[45] and *SRSF5* is a member of the serine/arginine (SR)-rich protein family that regulates pre-mRNA alternative splicing.[46] *SRSF5* has been shown to function as an oncogene in some human tumors, such as oral squamous cell carcinoma.[47] Additional studies are required to investigate the mechanism by which these fusions might stabilize their 5' partner expression, and the significance of this to development or progression of CD4$^+$ PTCL. *KLF10* expression was significantly increased in CD4$^+$ PTCLs expressing the *KLF10-YWHAZ* fusion both compared to other CD4$^+$ PTCLs and control samples (Figure 57). The predicted *KFL10-YWHAZ* fusion joins a truncated *KLF10* CDS to the 5' UTR of *YWHAZ*. Thus, protein-based assays like immunoblotting would be a useful next step to assess whether increased amounts of *KLF10* mRNA in *KLF10-YWHAZ*-expressing CD4$^+$ PTCLs are also associated with increased production of KLF10 protein, as translation may be disrupted by the fusion event.

Table 11: Differential expression of 5' and 3' partner genes in CD4$^+$ PTCLs positive for that fusion gene by PCR versus CD4$^+$ PTCLs that were negative for that fusion gene by PCR. Significant results are shown in bold.

Fusion gene	5' gene log2 fold change	5' gene padj	3' gene log2 fold change	3' gene padj
B2M-GNAS	**0.288**	**0.036**	0.0988	0.50
JPT1-GATD3A	0.034	0.863	-0.014	0.965
KLF10-YWHAZ	**1.371**	**2.02E-08**	0.110	0.473
MROH1-GATD3A	**0.610**	**0.012**	0.080	0.891
MYC-TRIB1	0.191	0.398	**0.854**	**4.00E-04**
NCL-PTMA	0.156	0.279	0.042	0.841
PER1-EIF5A	-1.091	0.212	-0.074	0.918
PTMA-HMGB1	0.011	0.973	-0.061	0.814
PTMA-LMO4	-0.020	0.948	0.263	0.658
REV3L-FYN	0.059	0.955	-0.073	0.946
SRSF5-LMO4	**0.776**	**6.98E-05**	0.144	0.716
TOX2-LMO4	0.032	0.957	0.313	0.449
TPD52L2-GATD3A	-0.071	0.717	0.032	0.917

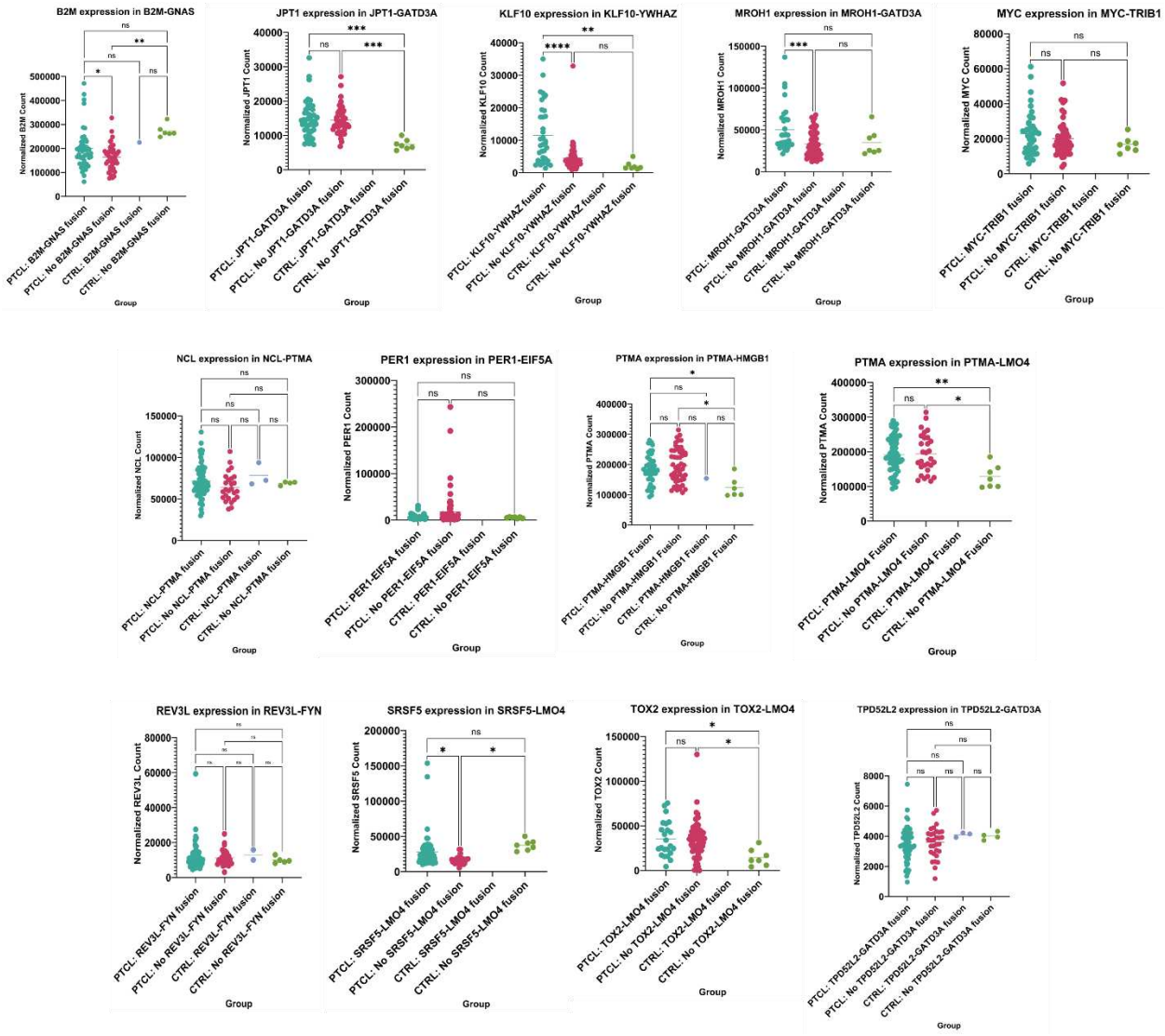


Figure 57: Differential expression of 5' fusion partner genes. For the *B2M-GNAS, KLF10-YWHAZ, MROH1-GATD3A,* and *SRSF5-LMO4* fusions, 5' partner gene expression was significantly higher in CD4$^+$ PTCLs that were PCR+ for that fusion gene compared to CD4$^+$ PTCLs that were negative by PCR, based on a Tukey's multiple comparisons test. For the *JPT1-GATD3A*, *PTMA-HMGB1, PTMA-LMO4* and *TOX2-LMO4* fusions, expression of the 5' partner gene was significantly higher in CD4$^+$ PTCL than in dogs without lymphoma, but not significantly different compared to CD4$^+$ PTCLs that lacked the fusion gene.

Regarding differential expression of 3' partner genes in our candidate fusion genes, only the *MYC-TRIB1* fusion gene was associated with significantly increased expression of its 3' fusion partner (*TRIB1*) compared to CD4$^+$ PTCLs without that fusion gene (Figure 58, Table 11). While *MYC-TRIB1* was detectable in some of our control samples by PCR, gene expression data was only available for control samples that were PCR-negative for *MYC-TRIB1. TRIB1* expression was also significantly increased in PTCLs that were PCR-positive for *MYC-TRIB1* compared to these control

samples. The *MYC-TRIB1* fusion is predicted to result in an out-of-frame sequence that is unlikely to result in functional protein production. Therefore, protein-based assays like immunoblotting would be a valuable next step to assess whether *TRIB1* overexpression in these *MYC-TRIB1*-expressing CD4$^+$ PTCLs corresponds to increased production of TRIB1 protein.

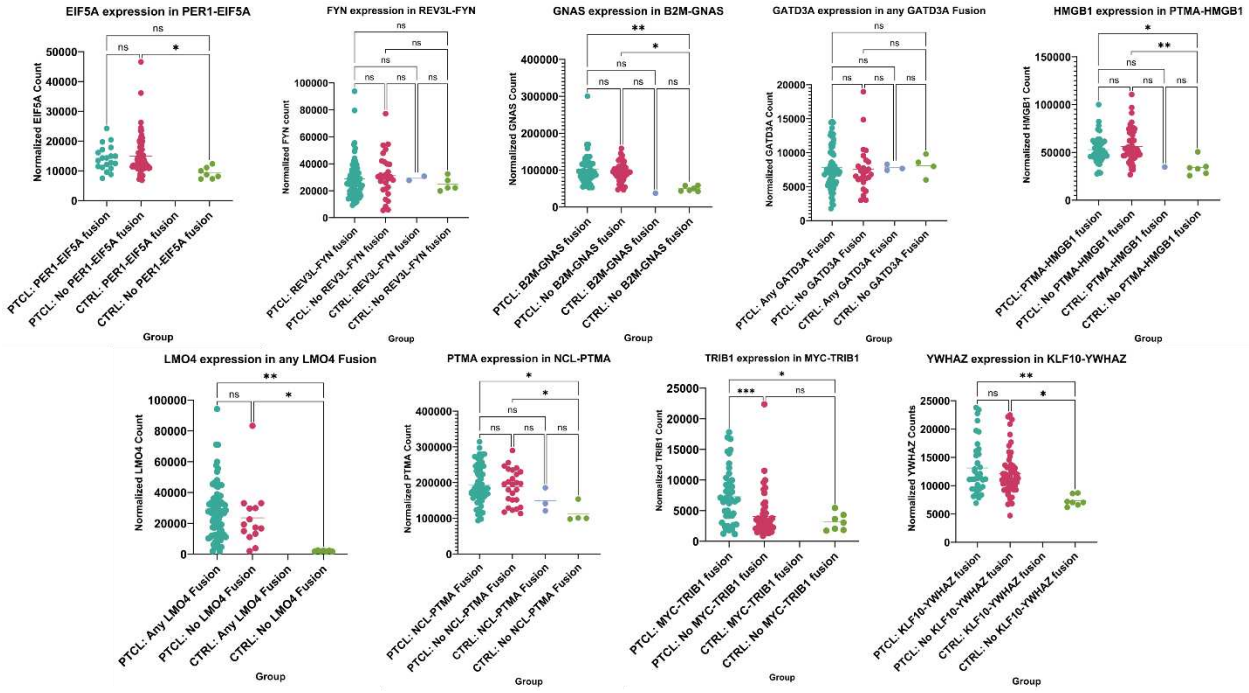


Figure 58: Differential expression of 3' fusion partner genes. For the *MYC-TRIB1* fusion gene, 3' partner gene expression was significantly higher in CD4$^+$ PTCLs that were PCR+ for that fusion gene compared to CD4$^+$ PTCLs that were negative by PCR, based on a Tukey's multiple comparisons test. For the *B2M-GNAS, PTMA-HMGB1, NCL-PTMA, KLF10-YWHAZ*, and any fusion involving *LMO4* as a 3' partner, expression of the 3' partner gene was significantly higher in CD4$^+$ PTCL than in dogs without lymphoma, but not significantly different compared to CD4$^+$ PTCLs that lacked the fusion gene.

Gene expression profiling offers insight into the impact of tumor-specific fusions *MROH1-GATD3A*, *TOX2-LMO4*, and *PER1-EIF5A* in CD4$^+$ PTCL

To further evaluate the 3 tumor-specific fusion genes revealed by our PCR assay, we conducted differential expression analysis comparing fusion-positive CD4$^+$ PTCLs (as determined by PCR) to fusion-negative CD4$^+$ PTCLs for the *MROH1-GATD3A, TOX2-LMO4*, and *PER1-EIF5A* fusions, respectively.

The *MROH1-GATD3A* fusion breakpoint predicted by FusionCatcher joins the coding sequences of *MROH1* and *GATD3A* in an in-frame manner, which could result in the translation of a novel fusion protein. The function of these genes as drivers of cancer is not well explored, but *GATD3A* was one of the most recurrent 3' fusion partners predicted in our RNA-seq data. While its function remains largely unknown, *MROH1* has been shown to be co-expressed with 33 other genes in breast cancer.[48] One of these genes, *MAPK15*, which has been previously implicated as an oncogene in some human cancers,[49] had significantly higher expression in *MROH1-GATD3A*-expressing CD4$^+$ PTCLs based on a pairwise comparison of normalized count means (Figure 59A). However, this gene was not found to be significantly differentially expressed based on analysis with DESeq2 (log2fc = 0.75, padj = 0.19). CD4$^+$ PTCLs that were PCR-positive for the *MROH1-GATD3A* fusion did not cluster discretely from CD4$^+$ PTCLs that were PCR-negative on a global gene expression level (Figure 59B).

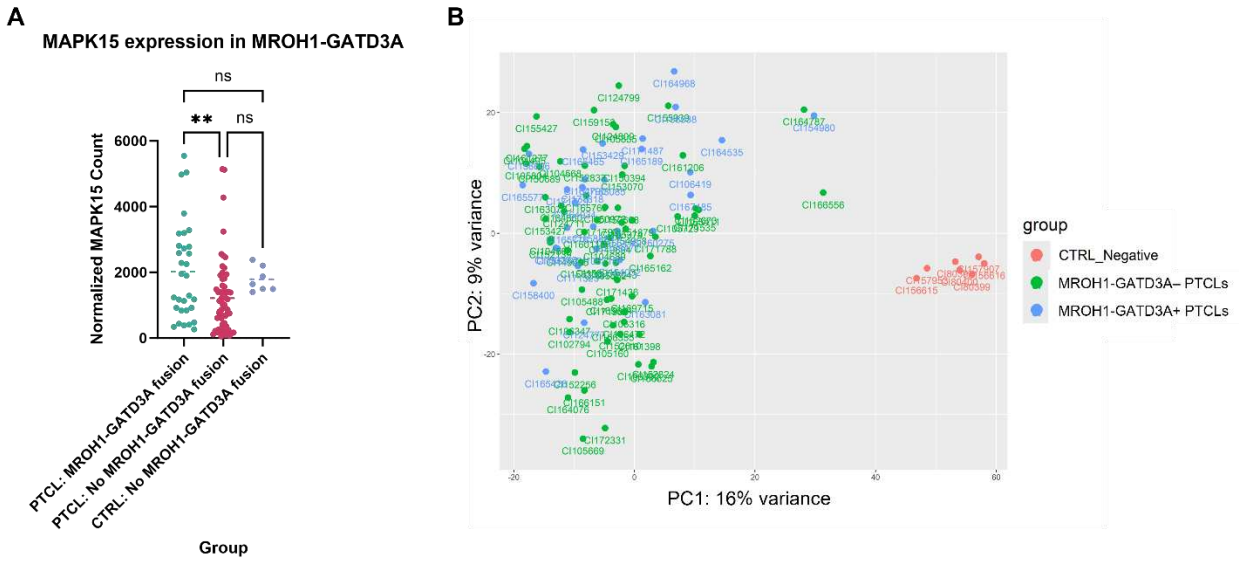


Figure 59: The gene expression profile of *MROH1-GATD3A*+ CD4$^+$ PTCLs compared to *MROH1-GATD3A*– *CD4$^+$ PTCLs*. (**A**) A Tukey's pairwise comparison test comparing the means of *MAPK15* normalized counts revealed significantly increased expression of *MAPK15*—a gene coexpressed with *MROH1* in human breast cancer—in *MROH1-GATD3A*-expressing CD4$^+$ PTCLs. (**B**) *MROH1-GATD3A*-expressing CD4$^+$ PTCLs (blue) do not have a distinctive global gene expression profile compared to *MROH1-GATD3A*– CD4$^+$ PTCLs (green), as illustrated by principal component analysis. Control CD4$^+$ nodal lymphocytes and CD4$^+$ thymocytes (red) cluster distinctly from tumor samples.

Overall, a relatively low number of genes were significantly up- or downregulated in *MROH1-GATD3A*-expressing CD4$^+$ PTCLs compared to CD4$^+$ PTCLs without this (Supplemental Table 21). One of the top upregulated genes in *MROH1-GATD3A*-expressing CD4$^+$ PTCLs compared to other CD4$^+$ PTCLs was *TGFB3* (log2fc = 1.97, padj = 2.22E-05), but TGF-β signaling gene signatures were not enriched in this subset by GSEA. Another one of the top upregulated genes in *MROH1-GATD3A*-expressing CD4$^+$ PTCLs was *NRG2* (log2fc = 1.25, padj = 0.0004), a ligand of the ERBB3 and ERBB4 receptors involved in activation of the PI3K oncogenic signaling pathways.[50] However, these CD4$^+$ PTCLs were not significantly enriched for ERBB or PI3K signaling associated gene signatures by GSEA compared to CD4$^+$ PTCLs lacking the *MROH1-GATD3A* fusion. This restriction of *MROH1-GATD3A*-associated changes in gene expression to individual genes without apparent changes in enrichment for their broader pathways makes the significance of these changes in CD4$^+$ PTCL unclear.

The *TOX2-LMO4* breakpoint predicted by FusionCatcher results in the formation of an in-frame sequence composed of the two coding sequences of each partner gene, which may be translated as a novel fusion protein. Expression of *TOX2* and *LMO4* are both higher in CD4$^+$ PTCL compared to normal CD4$^+$ nodal lymphocytes and CD4$^+$ thymocytes; however, their expression was not significantly different between CD4$^+$ PTCLs that were PCR-positive for *TOX2-LMO4* versus tumors that were negative (Figure 57, Figure 58). *TOX2* has been implicated in leukemogenesis of acute T-cell leukemia through repression of TIM3 (encoded by *HAVCR2*), a checkpoint inhibitor of T-cell activity.[51] *HAVCR2* was not differentially expressed in *TOX2-LMO4*+ CD4$^+$ PTCLs compared to *TOX2-LMO4*– CD4$^+$ PTCLs (log2fc = 0.16, padj = 0.84), although it was significantly upregulated in *TOX2-LMO4*+ CD4$^+$ PTCLs compared to control CD4$^+$ nodal lymphocytes and thymocytes (log2fc = 2.49, padj = 0.0005). *LMO4* is a member of the LIM-domain-only protein family, which function in the normal development of various cell types, including hematopoietic and neuronal cells.[52] All

members of this protein family have been associated with cancer development or progression.[52] Fusion genes involving either *LMO1* and *LMO2* as gene partners have been implicated in the development of human T-cell acute lymphoblastic leukemia (T-ALL),[52] and *LMO4* has been found to promote cancer cell invasion and proliferation by activating PI3K/AKT/mTOR signaling in gastric carcinoma cells and non-small cell lung cancer cells.[53,54]

TOX2-LMO4-expressing $CD4^+$ PTCLs were significantly enriched for TNF and NF-κB signaling compared to other $CD4^+$ PTCLs (NES = 1.75, padj = 0.005) (Figure 60A). Leading edge analysis revealed this enrichment was most significantly driven by increased expression of *IRS2* (log2fc = 1.55, padj = 0.0005), *TNFAIP3* (log2fc = 1.54, padj = 0.009), COX-2 enzyme encoding *PTGS2* (log2fc = 0.93, padj = 0.01), and NF-κB co-activator *KLF6*[55] (log2fc = 1.06, padj = 0.018). *IRS2* contributes to NF-κB activation through the PI3K/AKT pathway, and *LMO4* itself has also been shown to activate this pathway.[53,56] However, while PI3K/AKT signaling is generally enriched in $CD4^+$ PTCL (as demonstrated in Chapter One), this pathway was negatively enriched in *TOX2-LMO4*+ PTCLs compared to *TOX2-LMO4*– PTCLs in our study (NES = -1.81, padj = 0004) (Figure 60A). Interestingly, a number of cell cycle-associated pathways were negatively enriched in *TOX2-LMO4*+ PTCLs compared to *TOX2-LMO4*– PTCLs (Figure 60A), suggesting that these tumors may be less proliferative. This finding could be of clinical interest given the resistance of quiescent tumor cells to common chemotherapeutic agents that target cycling cells.

At a global gene expression level, $CD4^+$ PTCLs that expressed *TOX2-LMO4* by PCR did not cluster distinctly from other $CD4^+$ PTCLs (Figure 60B), suggesting that while these tumors may show greater enrichment for TNF and NF-κB signaling and decreased enrichment for cell cycle signatures, it does not result in a globally distinct molecular phenotype. The overexpression of genes along the NF-κB pathway, however, could represent a possible therapeutic target for tumors expressing this fusion gene. Additional studies investigating the dependency of *TOX2-LMO4*+ $CD4^+$

PTCLs on TNF and NF-κB signaling through the use of TNF and IKK inhibitors *in vitro* and/or *in vivo* would be a valuable next step based on the findings of this study.

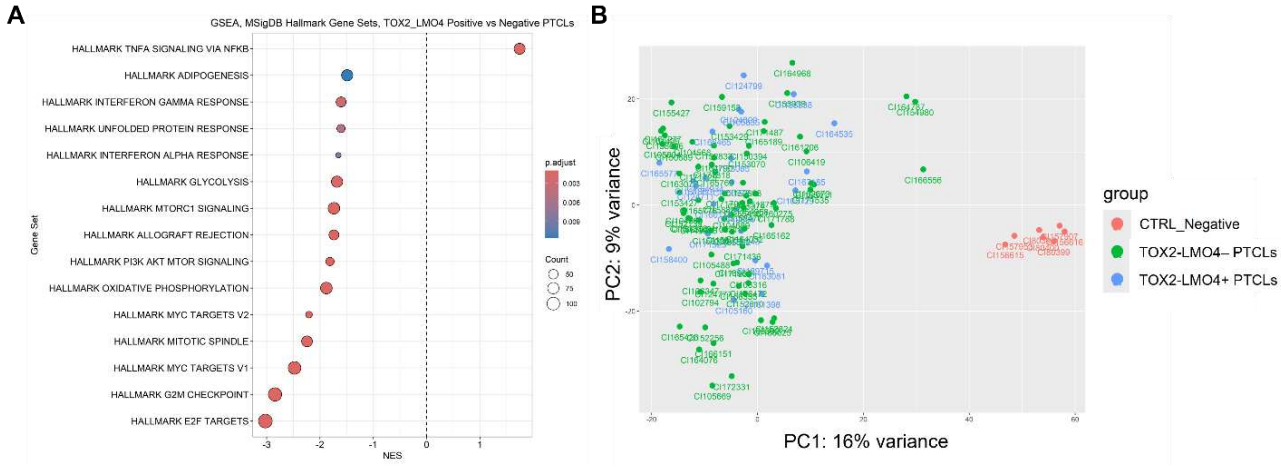


Figure 60: The gene expression profile of *TOX2-LMO4*+ CD4⁺ PTCLs compared to *TOX2-LMO4*– *CD4⁺ PTCLs*. (**A**) *TOX2-LMO4*-expressing CD4⁺ PTCLs showed significant enrichment for TNFA signaling via NF-kB by GSEA based on publicly available gene sets derived from the Molecular Signatures Database.[39] (**B**) *TOX2-LMO4*-expressing CD4⁺ PTCLs (blue) do not have a distinctive global gene expression profile compared to *TOX2-LMO4*– CD4⁺ PTCLs (green), as illustrated by principal component analysis. Control CD4⁺ nodal lymphocytes and CD4⁺ thymocytes (red) cluster distinctly from tumor samples.

The *PER1-EIF5A* breakpoint predicted by STAR-Fusion joins the first exon and part of the preceding 5' intron of *PER1* to the complete CDS of *EIF5A*. This fusion truncates the initial portion of the ORF preceding the CDS of *EIF5A,* so it is unclear whether EIF5A protein would still be translated. None of the *PER1* coding sequence is preserved in the fusion. In spite of this, neither *PER1* nor *EIF5A* mRNA expression was individually significantly impacted by the presence or absence of the fusion in CD4⁺ PTCL (Figure 57, Figure 58). However, assessment of the impact of this fusion on successful translation of the PER1 and EIF5A proteins requires the use of protein-based assays like immunoblotting. Inhibition of PER1 has been shown to confer resistance of human cancer cells to apoptosis,[57] so impaired translation of this protein in CD4⁺ PTCL could represent a mechanism by which tumor cells avoid apoptosis. EIF5A is an RNA-binding protein that functions in both initiation and elongation during translation, and its overexpression has been described as a negative prognosticator in several human cancers. Its tumor-promoting capabilities have been linked to its ability to regulate elongation of MYC and modulate expression of cytoskeletal-

associated proteins.[58,59] By GSEA, *PER1-EIF5A*-expressing CD4$^+$ PTCLs were enriched for the gene signature associated with epithelial-to-mesenchymal transition (Figure 61A), and the leading edge analysis revealed a variety of microenvironment-related genes as the top contributors to this enrichment, including metalloprotease-encoding *ADAM12*;[60] *LOX,* which encodes lysyl oxidase to trigger cross-linking and stabilization of extracellular matrix proteins;[61] and several collagen-encoding genes. These findings suggest that the *PER1-EIF5A* fusion gene may contribute to changes in the microenvironment that support tumor cell survival and proliferation, although additional studies are needed to ascertain the mechanism by which this fusion may be contributing to this process. The role of EIF5A in modulating the tumor microenvironment has been previously demonstrated in human pancreatic adenocarcinoma, where it was found to regulate expression levels of various cytoskeletal proteins to promote cancer cell migration.[59] Overall, a relatively low number of genes were significantly up- or downregulated in *PER1-EIF5A*-expressing CD4$^+$ PTCLs compared to CD4$^+$ PTCLs without this fusion (Supplemental Table 22). Among the top upregulated genes included *NPR3*, a gene known to promote proliferation and migration of human breast cancer cells,[62] and *ASCL2*, which induces cell proliferation and stemness in a variety of human cancer cell types and is associated with the creation of immunosuppressive tumor microenvironments.[63] At a global gene expression level, CD4$^+$ PTCLs that expressed *PER1-EIF5A* by PCR did not cluster distinctly from other CD4$^+$ PTCLs (Figure 61B).

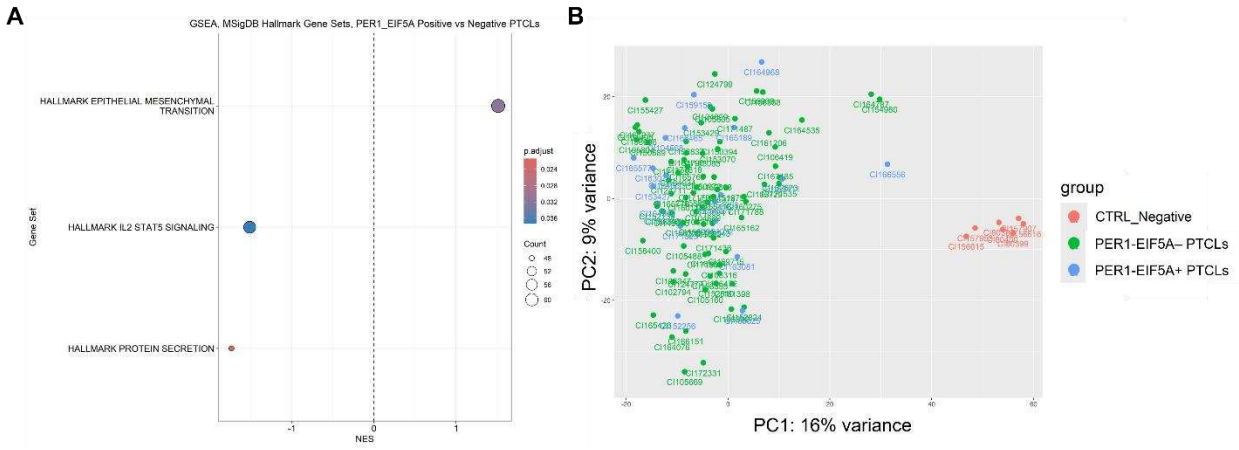


Figure 61: The gene expression profile of *PER1-EIF5A*+ CD4⁺ PTCLs compared to *PER1-EIF5A*– *CD4⁺ PTCLs*. (**A**) *PER1-EIF5A*-expressing CD4⁺ PTCLs showed significant enrichment for canonical genes involved in the epithelial-to-mesenchymal transition, based on publicly available gene sets derived from the Molecular Signatures Database.[39] (**B**) *PER1-EIF5A*-expressing CD4⁺ PTCLs (blue) do not have a distinctive global gene expression profile compared to *PER1-EIF5A*– CD4⁺ PTCLs (green), as illustrated by principal component analysis. Control CD4⁺ nodal lymphocytes and CD4⁺ thymocytes (red) cluster distinctly from tumor samples.

Taken together, these findings suggest that while the tumor-specific *MROH1-GATD3A, TOX2-LMO4,* and *PER1-EIF5A* fusions had a relatively unimpressive impact on the global gene expression profile of CD4⁺ PTCL, they still may contribute to the pathogenesis of these tumors by affecting expression of other cancer-associated genes in their regulatory network, although additional studies are required to elucidate the mechanism by which this may be occurring for each fusion. Additionally, the significance of these changes is unclear, given that we occasionally observed significantly altered expression of one gene in a signaling pathway without changes in the enrichment of these tumors for those pathways as a whole. Additional studies are needed to further explore the oncogenic impact—if any—of these fusion genes in CD4⁺ PTCL.

Long-read sequencing identified structural variants in canine CD4⁺ PTCL

Electrophoretic tracings of the distribution of input DNA fragment sizes for long-read sequencing of 6 CD4⁺ PTCL samples measured by TapeStation (Agilent Technologies, Inc., Santa Clara, CA) and Femto Pulse (Agilent Technologies, Inc., Santa Clara, CA) instruments are provided in Supplemental Figure 10. Approximately 3.4 million to 5.7 million HiFi reads were generated for

each sample by SMRT long-read sequencing for evaluation of structural variants (Table 12). The mean length of each HiFi read was 6,827 bp, with a range of 5,614-7,553 bp (Table 12). Median phred-like QV quality scores capturing the median probability of a correct base call across all reads ranged from Q37 to Q40 for our samples (Table 12), values which correspond to a base call accuracy of 99.98-99.99%.

Table 12: Quality metrics for HiFi long-read sequencing reads for 6 $CD4^+$ PTCL samples.

Sample	Barcode	Sex	Breed	Tissue	Age (yrs.)	Barcode Quality	HiFi Reads	HiFi Yield (GB)	HiFi Read Length (mean, bp)	HiFi Read Quality (median, QV)
CF78	bc2014	MC	SPRSP	Lymph node	6	97.1	3,735,101	28.0	7,507	Q37
CF79	bc2015	MC	HUS	Lymph node	4	96.9	3,474,221	23.1	6,660	Q38
CF72	bc2013	FS	OESD	Lymph node	4	96.9	5,733,070	32.2	5,614	Q40
CF128	bc2018	MC	BOX	Lymph node	7	97.4	3,766,001	25.0	6,639	Q39
CF89	bc2016	MC	MIX	Lymph node	8	96.9	3,833,417	26.8	6,989	Q37
CF90	bc2017	MC	SHTZ	Lymph node	9	97	3,715,223	28.0	7,553	Q37

Reads were aligned to both the CanFam3.1 and CanFam4 reference genomes for structural variant calling. The reference genome used did have an effect on structural variant calls. When aligned to the CanFam3.1 genome, a total of 126,672 structural variants were called across these 6 $CD4^+$ PTCL samples, with insertions representing the most common variant type at 71,502 (56%), followed by 47,161 (37%) deletions, 5,069 (4%) duplications, 2,590 (2%) chromosomal breaks, and 350 (0.3%) inversions (Figure 62A). When aligned to the CanFam4 genome, a total of 128,097 structural variants were identified across all samples. Insertions again made up the majority of these

variant calls at 69,842 (54%), followed by 49,665 (39%) deletions, 5,555 duplications (4%), 2,667 (2%) breakend variants, and 368 (0.3%) inversions (Figure 62B).

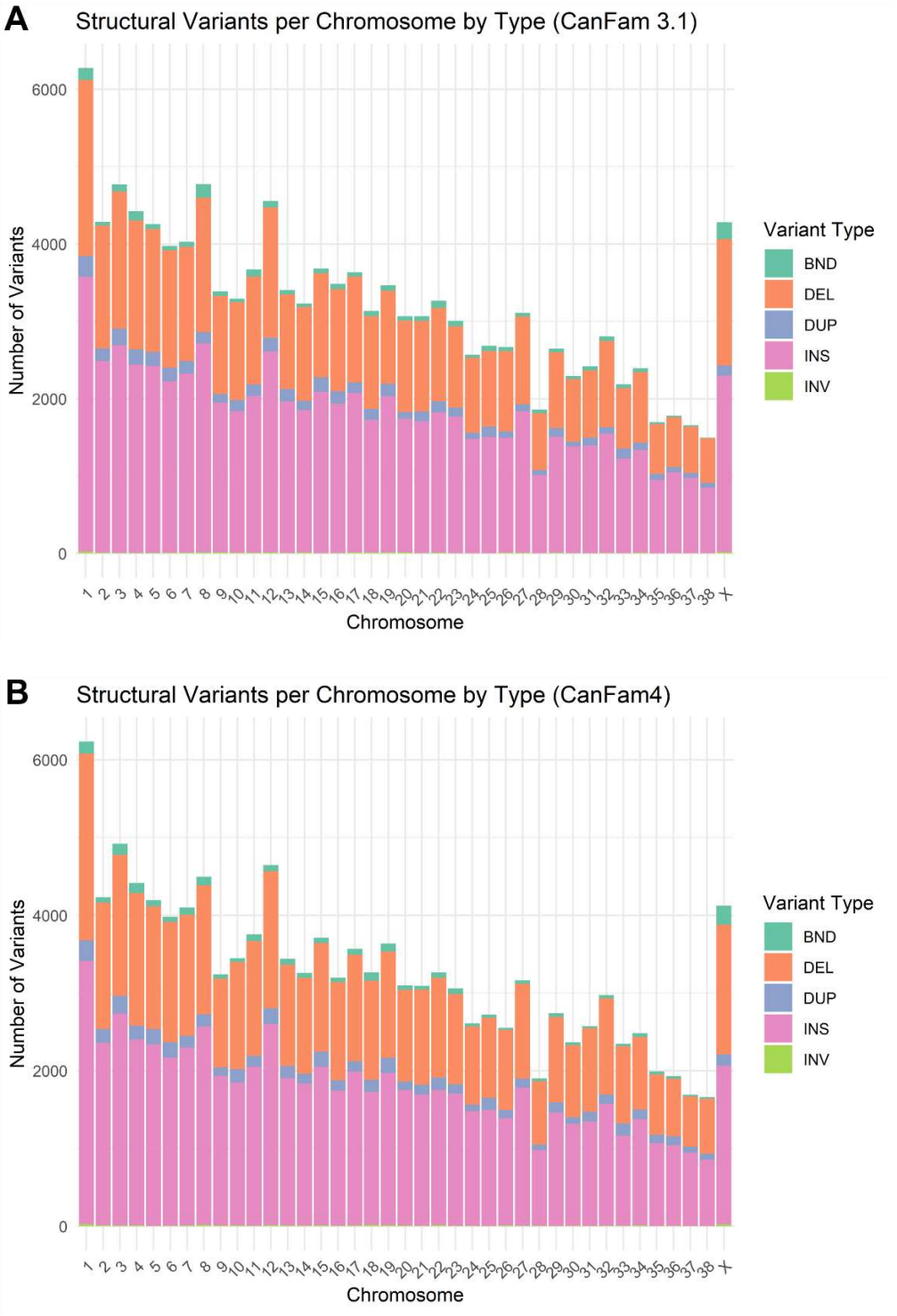


Figure 62: Summary of structural variant calls in CD4⁺ PTCL. Across all 6 samples, chromosome 1 had the most structural variant calls, and insertions were the most common structural variant type. Chromosomal breaks ("BND") were

most common in chromosomes 1, 8, and X when CanFam3.1 was used as the reference genome (**A**) and were most common in chromosomes 1, 3, and X when CanFam4 was used as the reference genome (**B**). BND = breakend. DEL = deletion. DUP = duplication. INS = insertion. INV = inversion.

Chromosome 1 had the highest number of total structural variant calls regardless of the reference genome used (Figure 62A-B). Chromosomes 1, 8, and X had the highest number of breakend variant calls based on alignment to CanFam3.1, with 156, 175, and 218 breakend variant calls, respectively (Figure 63A). Chromosomes 1, 3, and X had the highest number of breakend variant calls based on alignment to CanFam4, with 157, 142, and 246 breakend variant calls, respectively (Figure 62B, Figure 63).

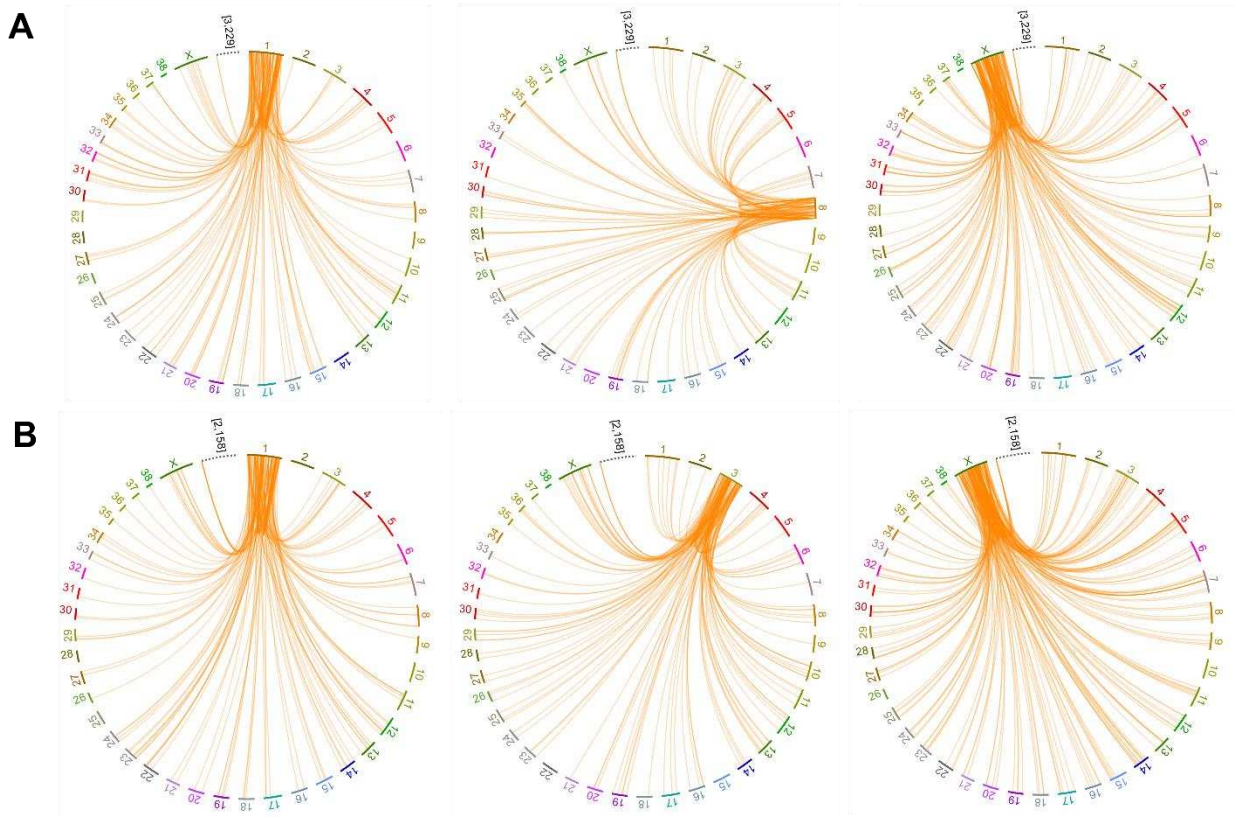


Figure 63: Distribution of translocation partners for the chromosomes with the most chromosomal breaks. (**A**) Chromosomes 1, 8, and X had the most breakend variant calls based on alignment to CanFam3.1. (**B**) Chromosomes 1, 3, and X had the most breakend variant calls based on alignment to CanFam4.

Circos plots illustrating the distribution of breakend variant calls in annotated genes for each sample are shown in Figure 64.

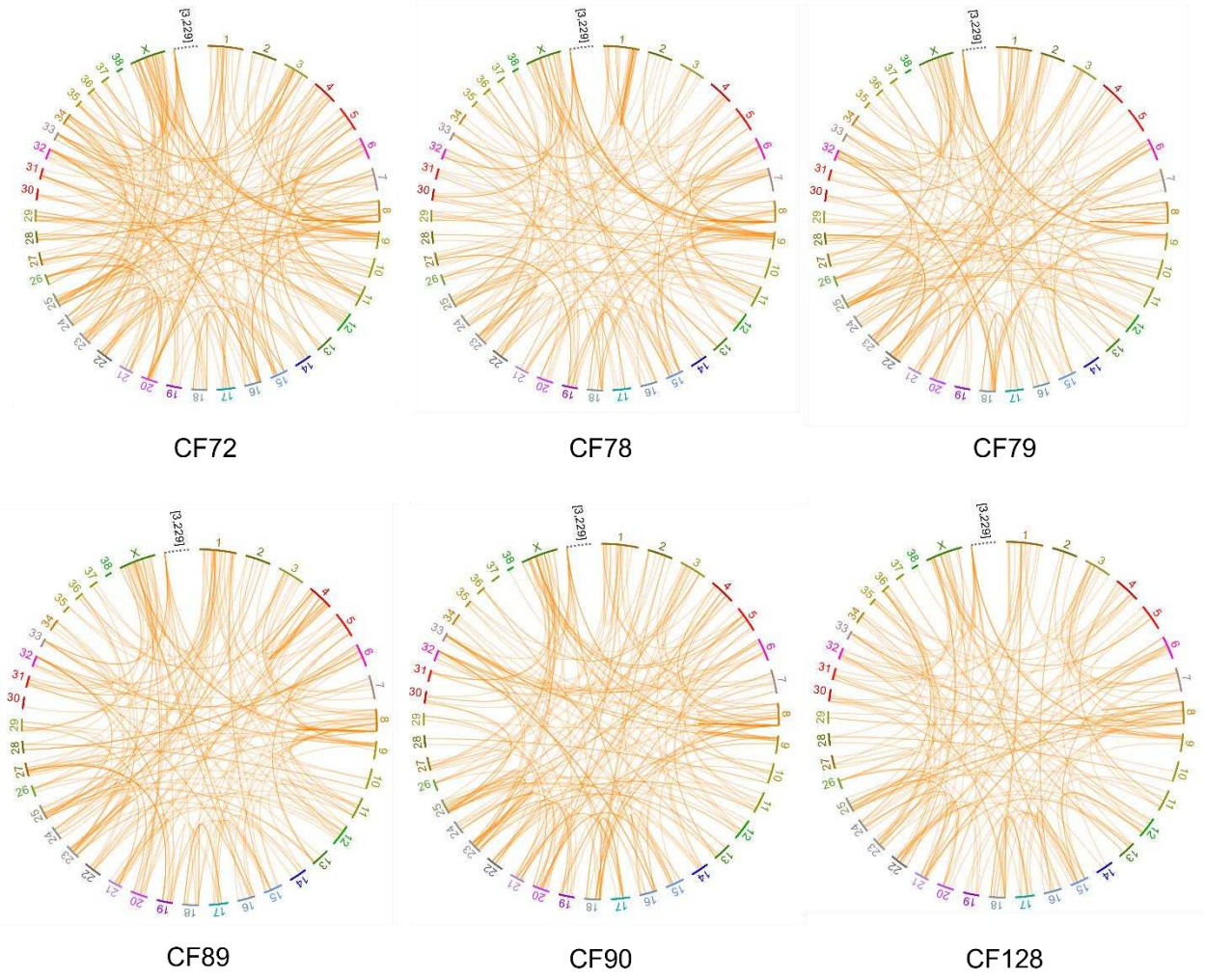


Figure 64: Translocation calls in CD4⁺ PTCL samples. Circos plots illustrate the connections between chromosomal breaks called in each CD4⁺ PTCL sample by long-read sequencing. Based on data using the CanFam3.1 reference genome and limited to chromosomal breaks implicating at least 1 annotated gene.

The total number of HNGC-annotated genes with variants called in both reference genomes was 5,071, and for 4,718 of these, the same variant type (insertion, deletion, breakend, etc.) was called in that gene. To further narrow down variants of interest that were present regardless of reference genome used, we considered only genes belonging to the OncoKB™ Cancer Gene List[64,65] and found that for 419 cancer-associated genes, the same structural variant type was identified in that gene regardless of reference genome used. Variants in cancer-associated genes that were identified in CD4⁺ PTCL samples in our study regardless of which reference genome was used for alignment included intronic deletions in *TGFBR1, TGFBR2,* and *TCF12* and intronic insertions and

deletions in *MSH2*. Deletions in *SYK* were also predicted with both genome alignments, although in CanFam3.1 they were confined to intronic regions while in CanFam4, one deletion affected the UTR (Figure 65A). An insertion/duplication of the coding sequence of the *TRAF5* gene was identified in one sample (CF89) regardless of reference genome used for alignment (Figure 65B-C).

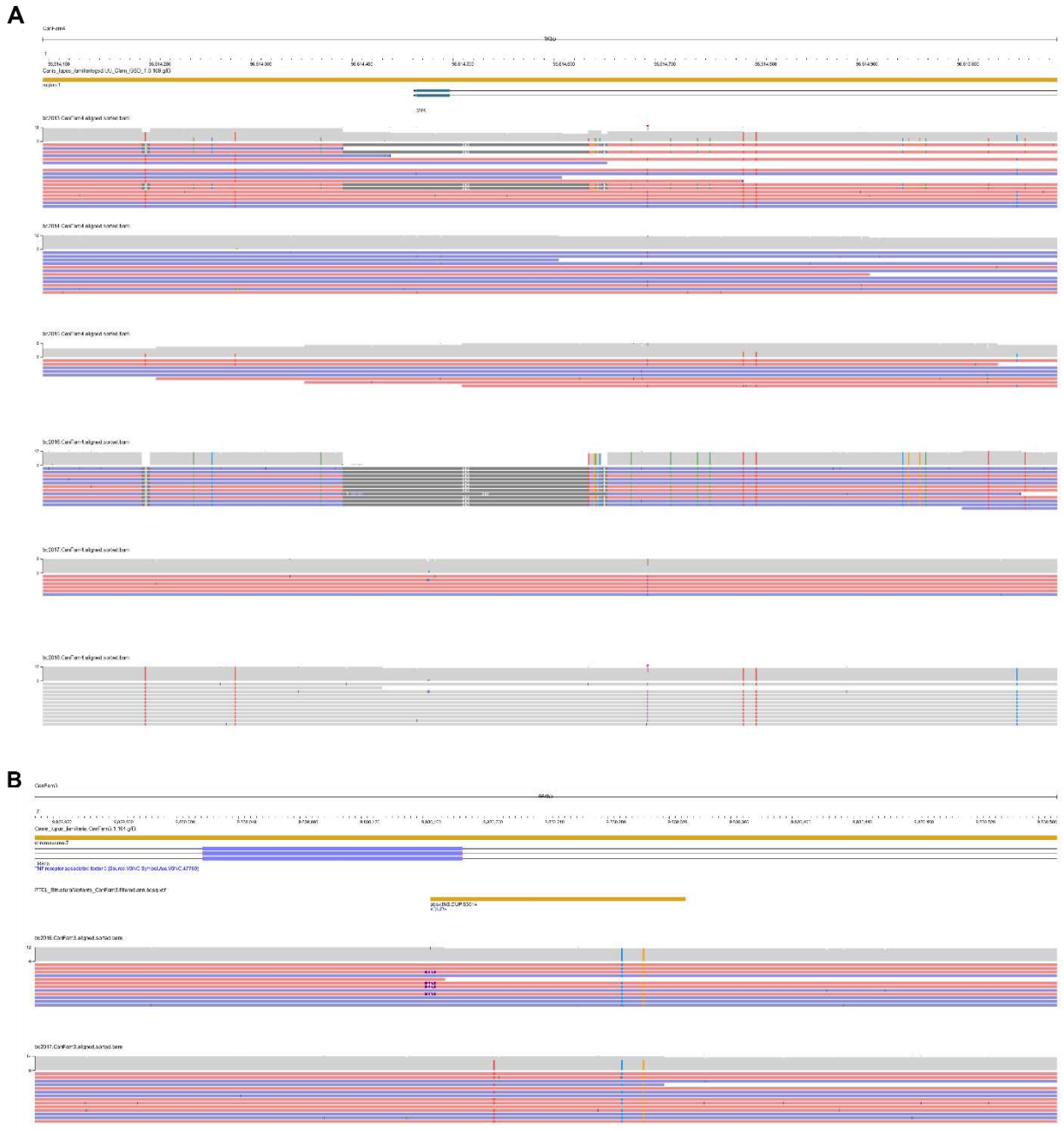


Figure 65: Structural variants in cancer-associated genes called in both the CanFam3.1 and CanFam4 alignments. (**A**) A deletion in the UTR of *SYK* was identified in CD4⁺ PTCL with the CanFam4 alignment. While *SYK* deletions were also a feature when the CanFam3.1 assembly was used for alignment, they all occurred in intronic regions. (**B-C**) An insertion/duplication of the coding sequence of the *TRAF5* gene was identified in one sample (CF89) regardless of whether the CanFam3.1 (**B**) or CanFam4 (**C**) reference genome was used for alignment. Reads aligned to the forward strand are shown in red and reads aligned to the reverse strand are shown in blue. Depth-of-coverage of reads at each position of the reference genome are shown in light gray above the aligned reads for each individual sample. Gene annotations are derived from the reference genome general feature format (GFF) file.

Long-read sequencing identified fusion events previously predicted from bulk RNA-seq data

Transcripts of several of the fusion genes predicted from bulk RNA-seq data were detectable in cDNA from this cohort by RT-PCR (Table 13). Despite this, none of these fusion gene candidates identified by our PCR assay were able to be identified in long-read DNA sequencing data from these samples.

Table 13: Fusion transcripts detected by PCR in the 6 CD4⁺ PTCLs of our long-read sequencing study.

Sample	Fusions detected by PCR
CF72	None
CF78	*JPT1-GATD3A, PTMA-HMGB1, PTMA-LMO4, REV3L-FYN*
CF79	*JPT1-GATD3A, NCL-PTMA, SRSF5-LMO4, TPD52L2-GATD3A, PER1-EIF5A, MYC-TRIB1, PTMA-LMO4, REV3L-FYN*
CF89	*JPT1-GATD3A, NCL-PTMA, PER1-EIF5A, PTMA-HMGB1, REV3L-FYN*
CF90	*B2M-GNAS, JPT1-GATD3A, MROH1-GATD3A, NCL-PTMA, SRSF5-LMO4, TPD52L2-GATD3A, PTMA-LMO4, REV3L-FYN*
CF128	*MROH1-GATD3A, SRSF5-LMO4, TPD52L2-GATD3A, MYC-TRIB1, PTMA-LMO4, REV3L-FYN, KLF10-YWHAZ*

However, there were 3 fusion events called in our long-read sequencing data that were also predicted by either FusionCatcher or STAR-Fusion from bulk RNA-seq data, although they were not incorporated as part of our PCR assay. These included *STAG2-SH2D1A*, *PLEKHA5-AEBP2*, and *SERPINB5-ENSCAFG00000031329* (Table 14).

Table 14: Fusion genes identified in CD4$^+$ PTCL by both PBSV in long-read sequencing data and by STAR-Fusion or FusionCatcher in bulk RNA-seq data. Genotype information for each sample is provided in the format of homozygous wild-type (0/0), homozygous mutant (1/1), or heterozygous mutant (0/1), followed by a colon, followed by the number of reads supporting the wild-type and mutant genotype, respectively, separated by a comma.

Fusion gene	Ref. genome	Breakpoint	CF72	CF78	CF79	CF89	CF90	CF128
STAG2-SH2D1A	CanFam 3.1	X:95452167-X:95740590	0/0:9,0	0/0:4,0	1/1:0,5	0/0: 5,0	0/0:8,0	0/0:4,0
STAG2-SH2D1A	CanFam 4	X:96214189-X:96502747	0/0: 9,0:9	0/0: 4,0	1/1: 0,5	0/0: 5,0	0/0:8,0	0/0:4,0
PLEKHA5-AEBP2	CanFam 4	27:18134347-27:18433068	0/0:13,0	0/0:1,0	1/1:0,5	0/0:9,0	0/0:25,0	0/0:13,0
SERPINB5-ENSCAFG00000031329	CanFam 3.1	1:13598001-1:13395961	0/1:0,1	1/1:0,2	0/1:0,1	0/1:3,3	0/1:0,1	1/1:0,4

The *STAG2-SH2D1A* fusion was called in the same dog (CF79) by both long-read sequencing and bulk RNA-seq data by STAR-Fusion (Figure 66). *STAG2* and *SH2D1A* are located approximately 200,000 bp apart on the forward strand of the X chromosome. *STAG2* is a tumor suppressor gene that encodes part of the cohesin complex, and loss of *STAG2* in some human cancers leads to alterations in chromatin structure, disruption of translational regulation, and impairment of DNA repair.[66] The *SH2D1A* gene encodes signaling lymphocyte activation molecule (SLAM)-associated protein (SAP), and mutation of this gene in human males leads to the development of X-linked lymphoproliferative disease.[67] While a fusion involving *SH2D1A* was only detected in 1 dog in our study, male dogs are overrepresented in canine CD4$^+$ PTCL, making the X chromosome a region of interest for uncovering possible driver mutations. In solid human tumors, *SH2D1A* can activate NF-

κB signaling to promote tumor cell growth and metastasis.[68]

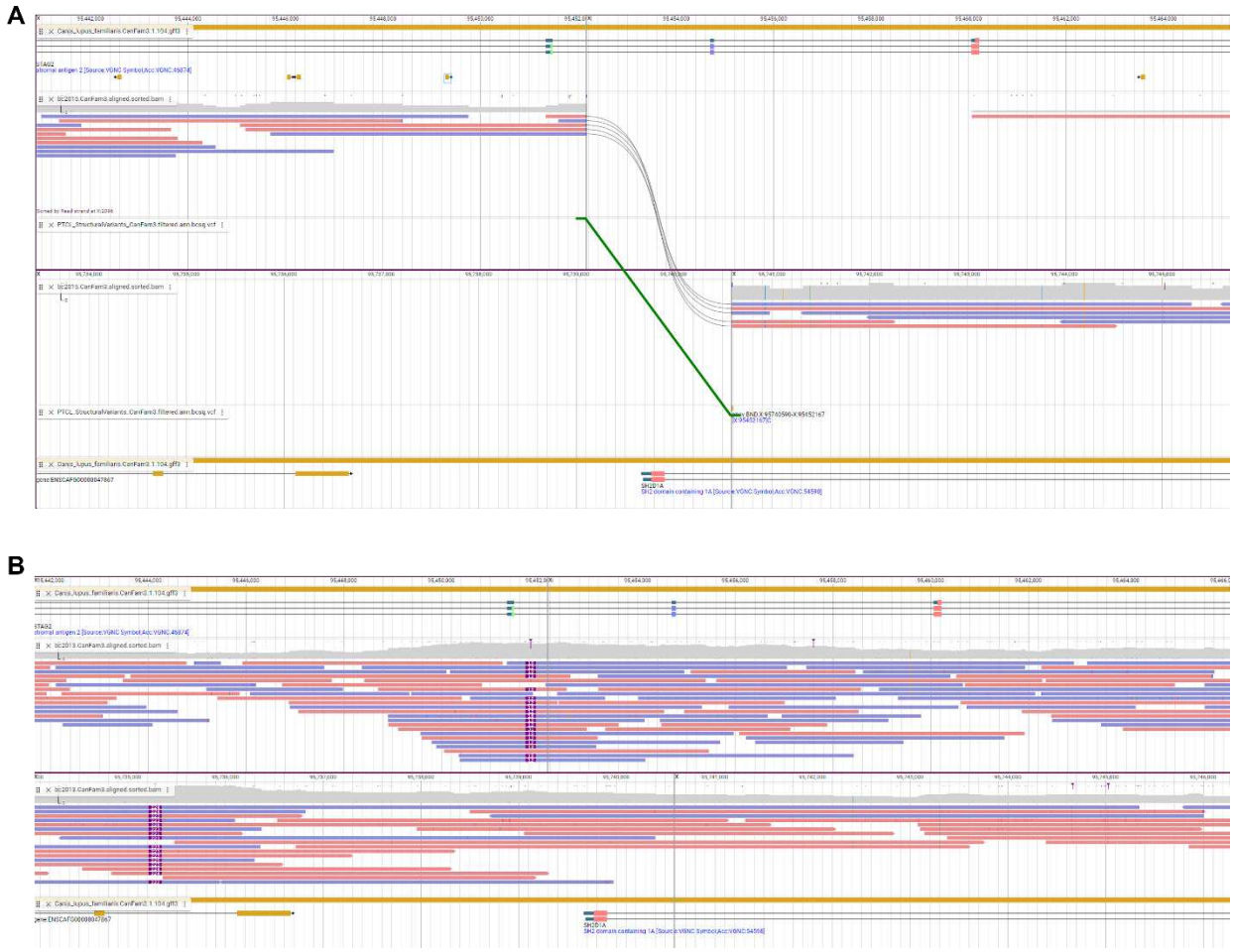


Figure 66: *STAG2-SH2D1A*. Visualization of the *STAG2-SH2D1A* in CD4⁺ PTCL sample CF79 (**A**) and the same region in a CD4⁺ PTCL without this fusion (CF72) (**B**). *STAG2* alignments are located in the top window of the split breakpoint view, while *SH2D1A* alignments are showcased in the bottom window. Reads aligned to the forward strand are shown in red and reads aligned to the reverse strand are shown in blue. Depth-of-coverage of reads at each position of the CanFam3.1 reference genome are shown in light gray above the aligned reads. Black curved lines connect reads that are split-aligned to the *STAG2* and *SH2D1A* genes on chromosome X. The green line is the variant call from the variant call file (VCF), with feet indicating directionality. Gene annotations are derived from the reference genome general feature format (GFF) file.

The *PLEKHA5-AEBP2* fusion was called in the same dog (CF79) by both long-read sequencing and bulk RNA-seq data by STAR-Fusion. *PLEKHA5* and *AEBP2* are located approximately 55,000 bp apart on the forward strand of chromosome 27. *PLEKHA5* plays an important role in cell signaling by recruiting proteins to specific membrane components,[69] and a few studies have shown that *PLEKHA5* can regulate tumor cell survival and proliferation.[69,70] *AEBP2* is a zinc finger protein that regulates transcription,[71] although its role in cancer is not well known.

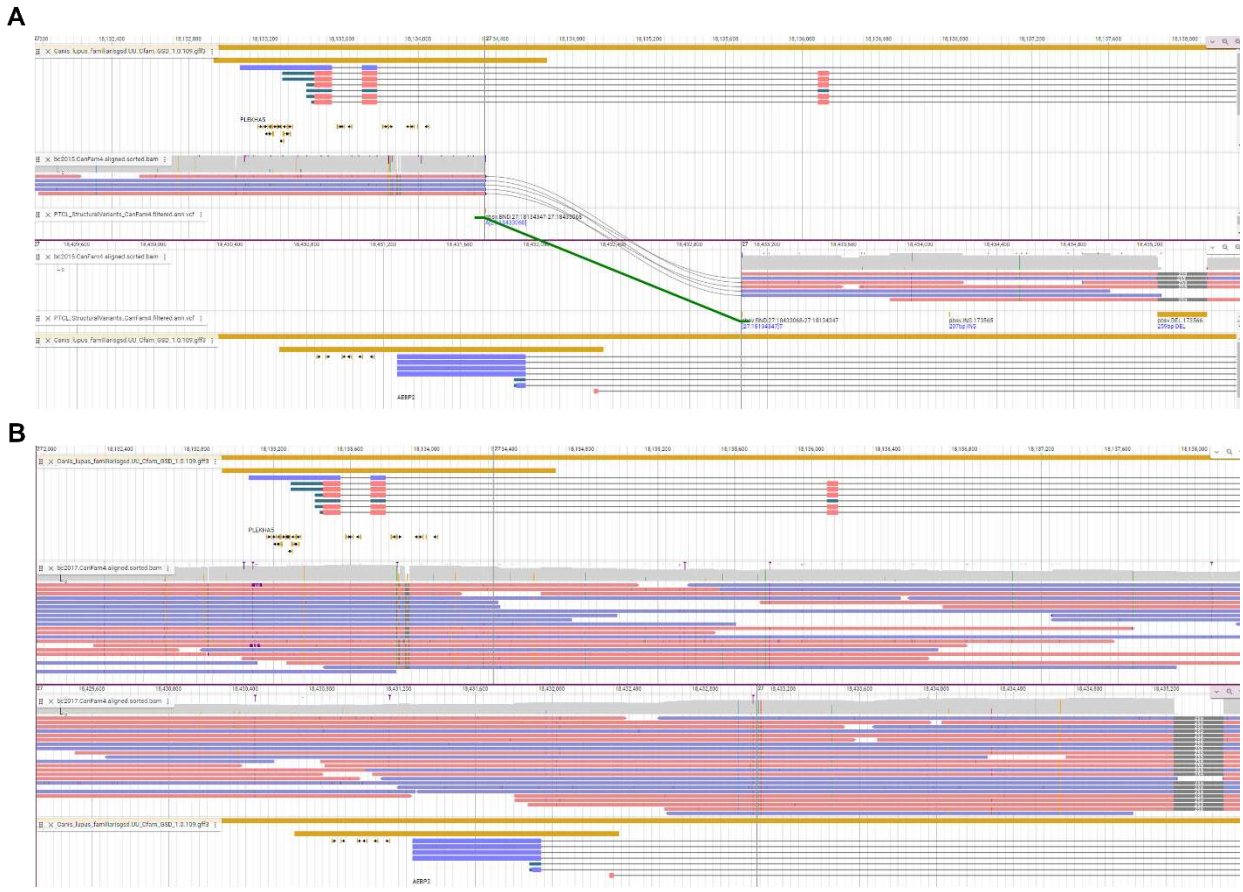


Figure 67: *PLEKHA5-AEBP2*. Visualization of the *PLEKHA5-AEBP2* in CD4$^+$ PTCL sample CF79 (**A**) and the same region in a CD4$^+$ PTCL without this fusion (CF90) (**B**). *PLEKHA5* alignments are located in the top window of the split breakpoint view, while *AEBP2* alignments are showcased in the bottom window. Reads aligned to the forward strand are shown in red and reads aligned to the reverse strand are shown in blue. Depth-of-coverage of reads at each position of the CanFam4 reference genome are shown in light gray above the aligned reads. Black curved lines connect reads that are split-aligned to the *PLEKHA5* and *AEBP2* genes on chromosome 27. The green line is the variant call from the variant call file (VCF), with feet indicating directionality. Gene annotations are derived from the reference genome general feature format (GFF) file.

A *SERPINB5-ENSCAFG00000031329* fusion was called in all 6 CD4$^+$ PTCLs in this study, although it was not discovered in these dogs by STAR-Fusion. However, it was called in 1 dog from another cohort by FusionCatcher. *SERPINB5* and *ENSCAFG00000031329* are located approximately 200,000 bp apart on the reverse strand of chromosome 1. *SERPINB5* is involved in a myriad of biological processes, and as such can function as both a tumor suppressor gene and an oncogene depending on the tumor type.[72] *ENSCAFG00000031329* encodes a long non-coding RNA (lncRNA). Whie lncRNAs do not encode proteins, they do function in regulation of gene expression, and their expression in cancer has been closely associated with chromosomal rearrangements.[2] Visualization

of the alignments for this fusion gene showed inconsistent alignments of split reads to multiple regions on the *ENSCAFG00000031329* gene, and these regions were accompanied by a large number of mismatches in the alignment to the reference genome (Figure 68). These findings were concerning for this call being an artifact rather than a true variant.

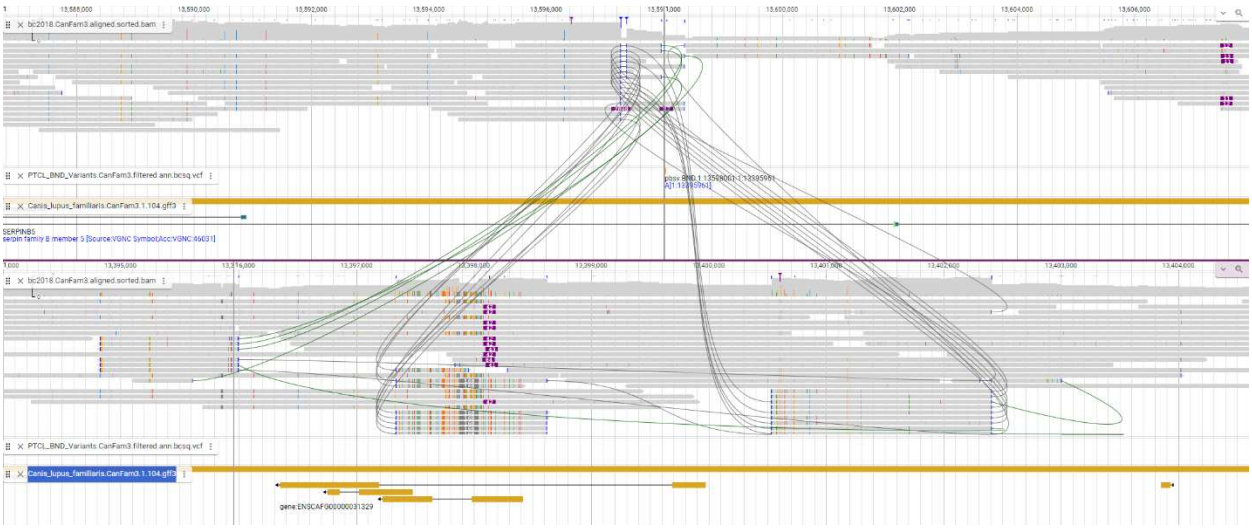


Figure 68: *SERPINB5-ENSCAFG00000031329*. This intrachromosomal translocation was predicted in 6/6 $CD4^+$ PTCLs and exhibited inconsistent alignments of split reads to multiple breakpoints on the *ENSCAFG00000031329* gene, with many alignment mismatches to the reference genome in this region as well. Reads mapped to the CanFam3.1 reference genome are shown in gray, with colored regions corresponding to mismatched bases. The larger purple rectangles correspond to insertions. Black and green curved lines connect reads that are split-aligned to the *SERPINB5* and *ENSCAFG00000031329* genes on chromosome 1. Gene annotations are derived from the reference genome general feature format (GFF) file.

Given the shared evidence of chromosomal aberrations detectable by long-read sequencing and production of chimeric mRNA transcripts detectable in RNA-seq data, screening of a larger cohort of tumor samples by PCR would be a valuable next step to further validate these fusions and assess their prevalence in $CD4^+$ PTCL.

<u>Long-read sequencing of $CD4^+$ PTCL identified structural variants in genes similarly affected in human PTCL-NOS</u>

Deletions in *TP53, CDKN2A,* and *PTEN* and gains or amplifications of *STAT3* and *MYC* are particularly characteristic of the GATA3-PTCL variant of PTCL-NOS, while gains in *BCL11B*, *BCL6, CD244, CD247, FASL6, TP63,* and *TPRG1* are features typical of the TBX21-PTCL subtype.[73]

Structural variant calling in long-read sequencing data from 6 canine CD4$^+$ PTCLs identified duplications in *STAT3* and *BCL11B* in 4/6 (67%) and 1/6 (17%) of cases, respectively. A 210- to 212-bp deletion in *TP53* in 3/6 (50%) cases (Figure 69), a 258-bp deletion in *CDKN1A* in 3/6 (50%) cases (Figure 70A), a smaller 57- to 60-bp *CDKN1A* deletion in all 6 samples (Figure 70B), and an 88- to 318-bp deletion in *CDKN2A* deletion in 4/6 (67%) cases (Figure 71) were also identified. These variants all arose in the intronic regions of these genes.

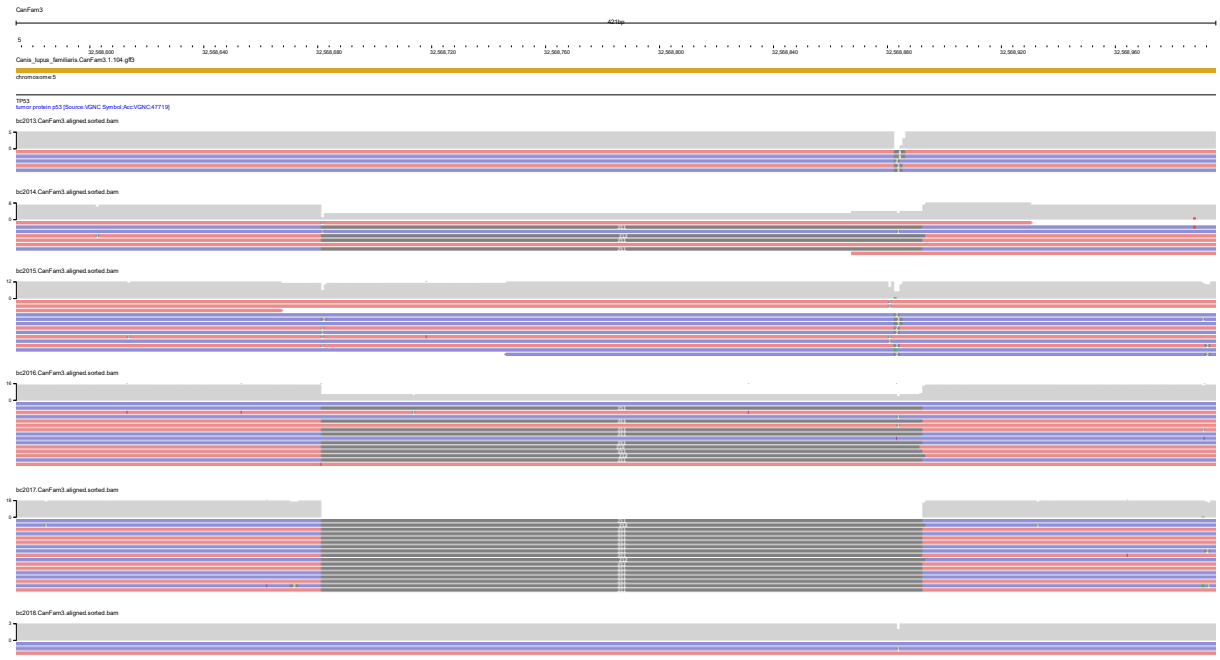


Figure 69: *TP53* deletion in CD4$^+$ PTCL. A 210- to 212-bp deletion in *TP53* (dark gray) was identified in all 3/6 (50%) of CD4$^+$ PTCL samples by long-read sequencing. Reads aligned to the forward strand are shown in red and reads aligned to the reverse strand are shown in blue. Depth-of-coverage of reads at each position of the CanFam3.1 reference genome are shown in light gray above the aligned reads for each individual sample. Gene annotations are derived from the reference genome general feature format (GFF) file.

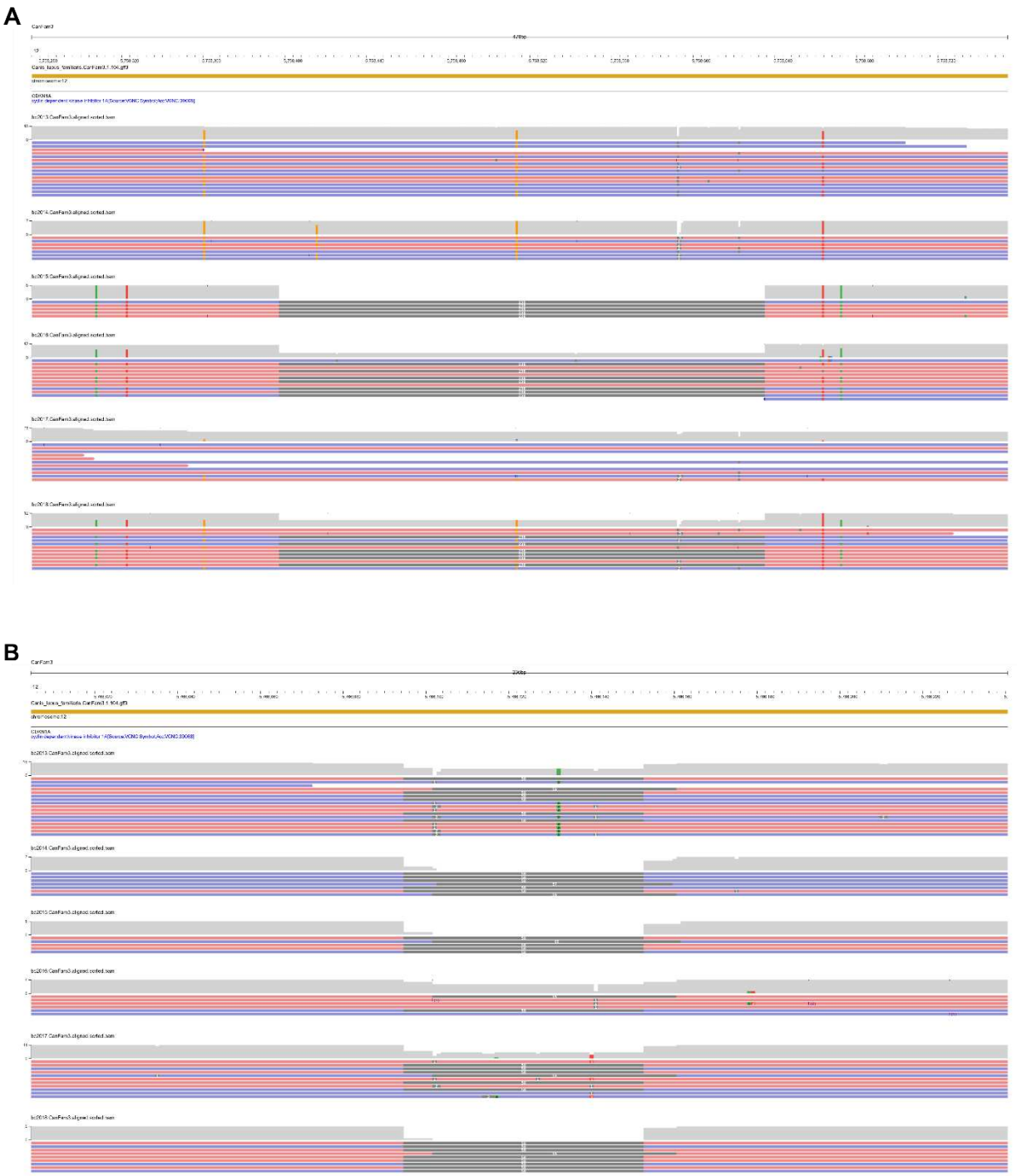


Figure 70: *CDKN1A* deletion in CD4⁺ PTCL. Two large deletions in *CDKN1A* (dark gray) were identified by long-read sequencing in CD4⁺ PTCL, including a 258-bp deletion in 3/6 (50%) of cases (**A**), and a smaller 57- to 60-bp deletion in all 6 samples (**B**). Reads aligned to the forward strand are shown in red and reads aligned to the reverse strand are shown in blue. Depth-of-coverage of reads at each position of the CanFam3.1 reference genome are shown in light gray above the aligned reads for each individual sample. Gene annotations are derived from the reference genome general feature format (GFF) file.

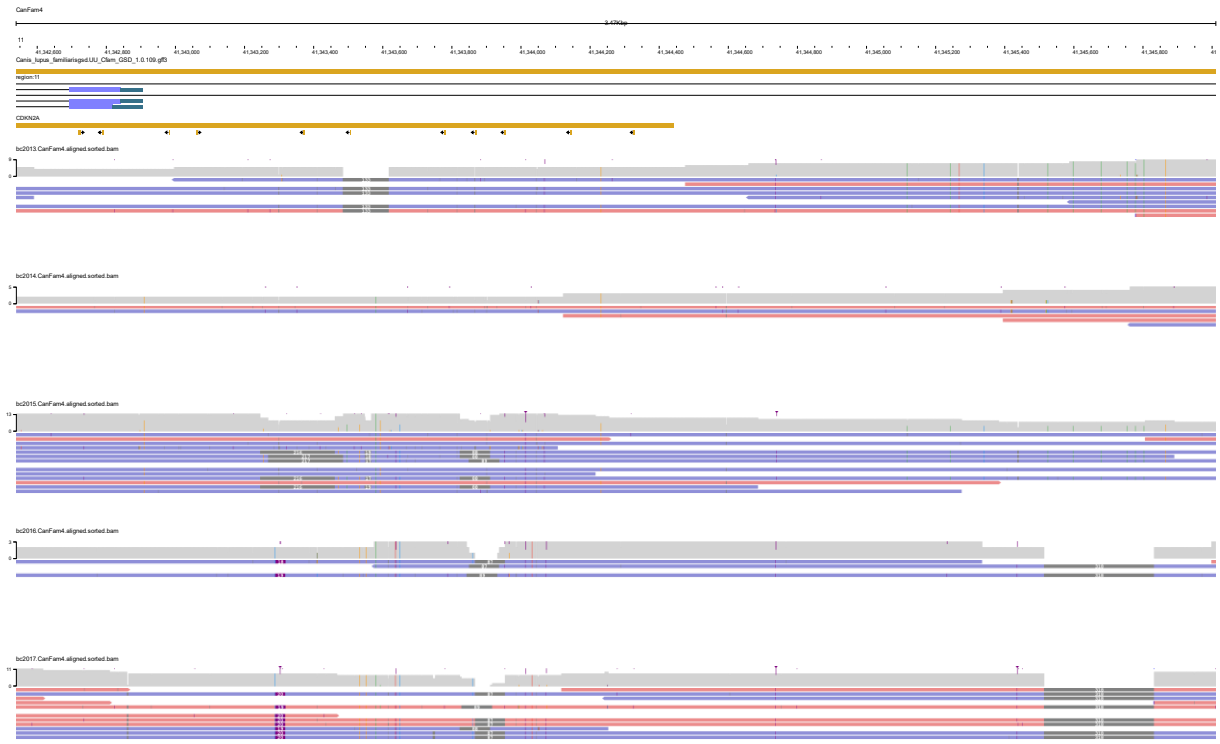


Figure 71: *CDKN2A* deletions in CD4⁺ PTCL. 88- to 318-bp deletions in *CDKN2A* (dark gray) were identified in 4/6 (67%) of CD4⁺ PTCL samples by long-read sequencing. Reads aligned to the forward strand are shown in red and reads aligned to the reverse strand are shown in blue. Depth-of-coverage of reads at each position of the CanFam3.1 reference genome are shown in light gray above the aligned reads for each individual sample. Gene annotations are derived from the reference genome general feature format (GFF) file.

An intronic 247-bp deletion in *PTEN* was detected in all 6 CD4⁺ PTCL samples in our study (Figure 72). Given that this deletion was present in all samples, arose in the same location, and was the same length, it is possible that this is simply a population variant with no pathologic significance. Future studies performing long-read sequencing on dogs without lymphoma would help determine if this *PTEN* deletion is truly a unique feature of oncogenic significance in CD4⁺ PTCL.

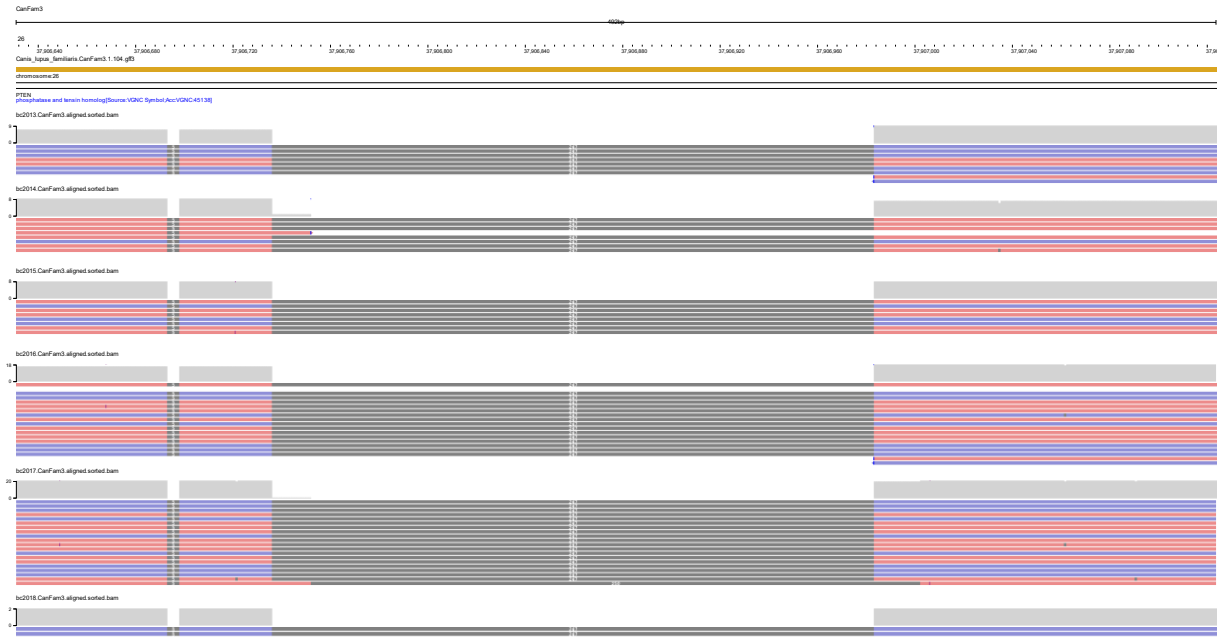


Figure 72: *PTEN* deletion in CD4⁺ PTCL. A 247- to 250-bp deletion in *PTEN* (dark gray) was identified in all 6 CD4⁺ PTCL samples by long-read sequencing. Reads aligned to the forward strand are shown in red and reads aligned to the reverse strand are shown in blue. Depth-of-coverage of reads at each position of the CanFam3.1 reference genome are shown in light gray above the aligned reads for each individual sample. Gene annotations are derived from the reference genome general feature format (GFF) file.

Discussion

In this study, we identified several recurrent fusion genes and other structural variants in canine CD4⁺ PTCL. Out of a selection of 27 fusion gene candidates predicted by fusion calling algorithms from bulk RNA-seq data, transcripts of 13 were able to be detected by RT-PCR in 120 samples of CD4⁺ PTCL. While fusion candidates discovered in normal dog samples using these fusion calling algorithms were filtered out prior to selection of candidates for validation by PCR, detectable chimeric transcripts were identified by PCR in dogs without lymphoma for 10/13 of the fusion genes in our PCR assay. Additionally, fusion genes predicted in only a small subset of CD4⁺ PTCLs from RNA-seq data were found to be much more recurrent in our PCR data. These findings indicate that fusion gene prediction from RNA-seq data may underestimate the true prevalence of these fusion transcripts, including in normal tissues, and rigorous validation by more sensitive methods like PCR is essential.

Fusion genes can contribute to oncogenesis through the generation of novel oncogenic chimeric proteins or changes in the regions responsible for regulating gene expression. In our study, 8/13 fusion genes confirmed by PCR—*REV3L-FYN, NCL-PTMA, PTMA-LMO4, TPD52L2-GATD3A, B2M-GNAS, JPT1-GATD3A, MROH1-GATD3A,* and *TOX2-LMO4*—resulted in an in-frame juxtaposition of the coding sequences of each partner gene that could result in the generation of novel fusion proteins, while 4/13 fusion genes confirmed by PCR—*KLF10-YWHAZ, PER1-EIF5A, PTMA-HMGB1,* and *SRSF5-LMO4*—resulted in juxtaposition of UTRs or intronic regions to truncated or complete coding sequences. In the case of *SRSF5-LMO4*, the 3' UTR was fused to the complete CDS of the *LMO4* gene with the ORF remained intact, which could place the expression of the *LMO4* gene under the control of the regulatory elements of the 3' UTR of *SRSF5*. However, while *LMO4* expression is significantly higher in $CD4^+$ PTCL compared to control nodal $CD4^+$ T cells, there was no appreciable difference in its expression between *SRSF5-LMO4*+ $CD4^+$ PTCLs compared to *SRSF5-LMO4*– $CD4^+$ PTCLs. The only fusion gene in our study where expression of the 3' partner gene was higher in fusion-positive versus fusion-negative PTCLs was *MYC-TRIB1*. The *MYC-TRIB1* fusion results in an out-of-frame sequence that is unlikely to result in a functional protein, so while *TRIB1* mRNA is overexpressed in *MYC-TRIB1*+ $CD4^+$ PTCLs, it is unclear if this would correspond to increased TRIB1 protein in tumor cells. Investigation of the protein-level impact of these fusions through techniques like immunoblotting is an important future direction for this study, both to investigate the potential for novel fusion protein generation from in-frame fusions and to determine the degree of impairment of protein production in fusion candidates expected to cause frameshifts or truncate coding sequences.

3/13 fusion genes detected by PCR— *MROH1-GATD3A, TOX2-LMO4*, and *PER1-EIF5A*— were present only in canine $CD4^+$ PTCL samples and not in dogs without lymphoma. Gene expression changes in $CD4^+$ PTCL related to the presence of a *MROH1-GATD3A* fusion were

relatively unremarkable. On the other hand, CD4$^+$ PTCLs that were PCR-positive for *TOX2-LMO4* showed greater enrichment for TNF and NF-κB signaling than CD4$^+$ PTCLs lacking this fusion, and CD4$^+$ PTCLs that were PCR-positive for *PER1-EIF5A* transcript exhibited a stronger tumor microenvironment signature. These findings could represent mechanisms by which these fusions contribute to the development or survival of CD4$^+$ PTCL cells and represent options for therapeutic targets. Future studies investigating the effect of agents like IκB kinase inhibitors in *TOX2-LMO4*+ CD4$^+$ PTCLs or tumor microenvironment modulators like thalidomide, lenalidomide, or pomalidomide in *PER1-EIF5A*+ CD4$^+$ PTCLs on the survival of these tumor cells *in vitro* and *in vivo* would offer additional insight into this potential. In human PTCL-NOS, oncogenic activation of NF-κB signaling is a feature of tumors expressing the *FYN-TRAF3IP2* fusion gene, and IκB kinase inhibitors have shown promising success in reducing tumor burden in mice with *FYN-TRAF3IP2*-induced lymphoma.[8]

The detection of fusion gene transcripts in both tumor and healthy tissues has been previously described. In the absence of an accompanying chromosomal aberration, these chimeric mRNAs are suspected to arise from alternative splicing events. Higher expression of fusion RNA in tumor samples than normal samples has been reported for some fusion genes in people, thought to be caused by clonal expansion of fusion-expressing cells that normally make up only a low proportion of the cells in a tissue.[34] When we compared the electrophoretic signal intensity of PCR-amplified fusion transcripts in our study, only the *REV3L-FYN* fusion was associated with a significantly higher signal intensity in PCR-positive CD4$^+$ PTCL samples compared to PCR-positive dogs without lymphoma. However, CE signal intensity is largely qualitative, and more sensitive quantitative methods for measuring analyte concentrations like qPCR should be considered in the future to more effectively assess any differences in fusion transcript expression between CD4$^+$ PTCL and non-neoplastic samples.

For some human gene fusions, it has been demonstrated that identical fusion RNA can be found in both neoplastic cells and their normal cell counterpart, but only in malignantly transformed cells is a concurrent chromosomal aberration detectable.[16,17] Studies by Gupta et al.[18,19] demonstrated that transfection of fusion RNAs into neoplastic cells lacking a chromosomal aberration can successfully induce the associated gene fusion at the DNA level, with associated transcription of additional fusion mRNA. Long-read sequencing to identify structural variants at the DNA level in $CD4^+$ PTCL did not identify an associated chromosomal aberration for the fusion gene transcripts confirmed by our PCR assay. However, it did identify 3 fusions—*STAG2-SH2D1A, PLEKHA5-AEBP2,* and *SERPINB5-ENSCAFG00000031329*—that were also among the predictions output by fusion calling algorithms from bulk RNA-seq data.

The *STAG2-SH2D1A* fusion, identified in 1 $CD4^+$ PTCL sample by both long-read sequencing and bulk RNA-seq, is an intrachromosomal fusion on the forward strand of the X chromosome. This candidate was of particular interest given the overrepresentation of male dogs in $CD4^+$ PTCL, which could suggest an X-linked genetic contribution to their predisposition for this neoplasm. The *SH2D1A* gene encodes signaling lymphocyte activation molecule (SLAM)-associated protein (SAP), and in human males, mutation of this gene causes X-linked lymphoproliferative disease.[67] In other human tumors, *SH2D1A* can activate NF-kB signaling to promote tumor cell growth and metastasis.[68] The *STAG2* gene is a tumor suppressor gene that plays a role in regulating translation and DNA repair,[66] so a fusion event that resulted in impairment of STAG2 translation could also potentially contribute to the pathogenesis of $CD4^+$ PTCL. *PLEKHA5-AEBP2*, resulting from an intrachromosomal rearrangement on the forward strand of chromosome 27, was another fusion mutually called in 1 $CD4^+$ PTCL sample by long-read sequencing and bulk RNA-seq. While the role of the zinc finger protein-encoding *AEBP2* in cancer is not well known, *PLEKHA5* has been shown to regulate tumor cell survival and proliferation through its role in recruiting effector proteins to

membrane components as part of cell signaling.[69,70] *SERPINB5-ENSCAFG00000031329*, resulting from an intrachromosomal rearrangement on the reverse strand of chromosome 1, was the third fusion discovered in $CD4^+$ PTCL both by long-read DNA sequencing and bulk RNA-seq. However, visualization of supporting read alignments revealed inconsistent alignment of reads to multiple breakpoints—even within the same dog—and a large number of mismatched base alignments to the reference genome in the implicated region of *ENSCAFG00000031329,* raising concern for an artifact rather than a true fusion. Given the evidence of concurrent chromosomal aberrations and chimeric mRNA transcripts detectable in RNA-seq data for these 3 fusion gene candidates, screening of a larger cohort of tumor samples by PCR would be a valuable next step to further validate these fusions and assess their prevalence in $CD4^+$ PTCL. Long-read sequencing also identified numerous structural variants—including fusion genes—in cancer-associated genes that, while not predicted from bulk RNA-seq data, may still be of interest to investigate more extensively in $CD4^+$ PTCL.

Aside from chromosomal breaks and fusion genes, long-read sequencing revealed other large structural variants in canine $CD4^+$ PTCL that implicated genes commonly mutated in human PTCL-NOS. The GATA3-PTCL variant of human PTCL-NOS is particularly prone to mutations affecting the *CDKN2A/TP53* axis, loss-of-function mutations and deletions in *PTEN*, and gains or amplifications of *STAT3* and *MYC*.[73–75] Gains in *BCL11B*, *BCL6, CD244, CD247, FASL6, TP63,* and *TPRG1* are more common in the TBX21-PTCL subtype of PTCL-NOS.[73] Large deletions in *PTEN, CDKN2A, TP53,* and *CDKN1A* were identified in 100%, 67%, 50%, and 50% of the 6 canine $CD4^+$ PTCLs in our study, respectively, as well as duplications of *STAT3* in 67% and duplications of *BCL11B* in 17%. These variants were all identified in intronic regions. While introns do not contain the protein-coding sequences of the gene, mutations affecting introns can impair normal splicing, affecting gene expression or disrupting the reading frame of the resulting transcript, resulting in a nonfunctional protein. These genes were not found to be differentially expressed in $CD4^+$ PTCLs

compared to normal nodal CD4$^+$ lymphocytes in bulk RNA-seq data, with the exception of *PTEN* in our smaller cohort of 25 CD4$^+$ PTCLs (described as Cohort 1 in previous chapters), which was downregulated (log2fc = -0.74, padj = 0.03). Protein-based assays like immunoblotting would be a valuable next step to assess how functional expression of the proteins encoded by these genes may be impacted by these variants.

Reference genome variability made interpreting the significance of some of these structural variants challenging. Structural variant discovery results were different based on the reference genome used for HiFi read alignment. For example, large deletions in *TP53* were only identified in CD4$^+$ PTCL samples using the CanFam3.1 reference genome, while deletions in *CDKN2A* were only identified using the CanFam4 reference genome. Disparities between reference genomes is an issue in human variant calling as well: one study found that approximately 5% of detected variants could not be converted between GRCh38 and GRCh37.[76] The CanFam4 reference genome improved upon gaps present in the CanFam3.1 assembly, 19.6% of which occurred within gene bodies.[77] While CanFam3.1 was generated using a 2005 7.4× Sanger sequencing framework, the CanFam4 reference genome was generated using a combination of Pacific Biosciences (PacBio) long-read sequencing, 10x Genomics Chromium Linked Reads, and HiC proximity ligation, and annotated with novel and published whole-genome sequence, assay for transposase-accessible chromatin (ATAC) sequencing, and RNA-seq.[77] A liftover of CanFam3.1 gap regions showed that 97% contained uniquely anchored sequences in CanFam4.[77] The mean mapping rate of HiFi reads across all long-read sequencing samples in our study was 99.949% for the CanFam3.1 alignment and 99.975% for the CanFam4 alignment (Supplemental Table 23), a difference that was statistically significant by a basic unpaired t-test (p = 0.02). This significantly greater genome coverage achieved when HiFi reads were aligned to the CanFam4 reference genome may reflect those efforts to address gap regions in CanFam3.1.

The more accurate structure of the CanFam4 reference assembly may justify higher confidence in structural variants identified using this genome alignment.

One limitation of our study was the lack of long-read sequencing data in samples from dogs without lymphoma. Some of the structural variants identified in our study, such as the previously described deletions in *PTEN*, were present in all 6 $CD4^+$ PTCL samples and were relatively uniform in length and location, raising concerns that these could represent population variants or inaccuracies in the reference genome assembly rather than a true mutation with pathologic potential. The inclusion of normal dogs could help clarify true variants from these artifacts. Additionally, the DNA used for long-read sequencing in our study was extracted using a protocol that produces predominantly 20-30 kb fragments, while recommended DNA extraction protocols for SMRT long-read sequencing typically result in predominantly 50-300+ kb fragments. Additional long-read sequencing experiments where DNA is extracted by the recommended techniques and normal dogs are included will help bolster the preliminary structural variant data gathered by this initial study.

Conclusions

Recurrent fusion genes are a feature of canine $CD4^+$ PTCL that may result from a combination of chromosomal aberrations and alternative splicing events. Some of the gene fusions identified in $CD4^+$ PTCL were accompanied by changes in gene expression and enrichment of these tumors for particular signaling pathways or TME elements, suggesting potential mechanisms by which these fusions may contribute to lymphomagenesis and possible avenues to explore for targeted therapies in tumors expressing these fusions. For a subset of the fusions predicted from bulk RNA-seq data of $CD4^+$ PTCL, fusion transcripts were also able to be detected by PCR in dogs without lymphoma, emphasizing the need for rigorous validation of fusion gene candidates discovered through high-throughput sequencing experiments. Future studies investigating the

protein-level impact of these fusion events, as well as other large structural variants identified in $CD4^+$ PTCL, are an important next step toward uncovering the role these mutations may play in the pathogenesis of these tumors and the opportunity they may offer as therapeutic targets.

REFERENCES

1. Mertens, F., Johansson, B., Fioretos, T. & Mitelman, F. The emerging complexity of gene fusions in cancer. *Nature Reviews Cancer* 15, 371–381 (2015).

2. Taniue, K. & Akimitsu, N. Fusion Genes and RNAs in Cancer Development. *Noncoding RNA* 7, (2021).

3. Ortiz de Mendíbil, I., Vizmanos, J. L. & Novo, F. J. Signatures of Selection in Fusion Transcripts Resulting From Chromosomal Translocations in Human Cancer. *PLOS ONE* 4, e4805 (2009).

4. Streubel, B., Vinatzer, U., Willheim, M., Raderer, M. & Chott, A. Novel t(5;9)(q33;q22) fuses ITK to SYK in unspecified peripheral T-cell lymphoma. *Leukemia* 20, 313–318 (2006).

5. Pechloff, K. *et al.* The fusion kinase ITK-SYK mimics a T cell receptor signal and drives oncogenesis in conditional mouse models of peripheral T cell lymphoma. *J Exp Med* 207, 1031–1044 (2010).

6. Dierks, C. *et al.* The ITK-SYK Fusion Oncogene Induces a T-Cell Lymphoproliferative Disease in Mice Mimicking Human Disease. *Cancer Research* 70, 6193–6204 (2010).

7. Debackere, K. *et al.* Fusion transcripts FYN-TRAF3IP2 and KHDRBS1-LCK hijack T cell receptor signaling in peripheral T-cell lymphoma, not otherwise specified. *Nature Communications* 12, 3705 (2021).

8. Moon, C. S. *et al.* FYN–TRAF3IP2 induces NF-κB signaling-driven peripheral T-cell lymphoma. *Nature Cancer* 2, 98–113 (2021).

9. Boddicker, R. L. *et al.* Integrated mate-pair and RNA sequencing identifies novel, targetable gene fusions in peripheral T-cell lymphoma. *Blood* 128, 1234–1245 (2016).

10. Molica, M., Abruzzese, E. & Breccia, M. Prognostic Significance of Transcript-Type BCR - ABL1 in Chronic Myeloid Leukemia. *Mediterr J Hematol Infect Dis* 12, e2020062 (2020).

11. Smith, C. *et al.* TAL1 activation in T-cell acute lymphoblastic leukemia: a novel oncogenic 3' neo-enhancer. *haematol* 108, 1259–1271 (2023).

12. Lei, Y., Lei, Y., Shi, X. & Wang, J. EML4-ALK fusion gene in non-small cell lung cancer (Review). *Oncol Lett* 24, 277 (2022).

13. Qin, F., Zhang, Y., Liu, J. & Li, H. SLC45A3-ELK4 functions as a long non-coding chimeric RNA. *Cancer Letters* 404, 53–61 (2017).

14. Guerra, E. *et al.* A Bicistronic CYCLIN D1-TROP2 mRNA Chimera Demonstrates a Novel Oncogenic Mechanism in Human Cancer. *Cancer Research* 68, 8113–8121 (2008).

15. Babiceanu, M. *et al.* Recurrent chimeric fusion RNAs in non-cancer tissues and cells. *Nucleic Acids Res* 44, 2859–2872 (2016).

16. Li, H., Wang, J., Mor, G. & Sklar, J. A Neoplastic Gene Fusion Mimics Trans-Splicing of RNAs in Normal Human Cells. *Science* 321, 1357–1361 (2008).

17. Yuan, H. *et al.* A Chimeric RNA Characteristic of Rhabdomyosarcoma in Normal Myogenesis Process. *Cancer Discovery* 3, 1394–1403 (2013).

18. Gupta, S. K., Jea, J. D.-Y. & Yen, L. RNA-driven JAZF1-SUZ12 gene fusion in human endometrial stromal cells. *PLOS Genetics* 17, e1009985 (2021).

19. Gupta, S. K., Luo, L. & Yen, L. RNA-mediated gene fusion in mammalian cells. *Proceedings of the National Academy of Sciences* 115, E12295–E12304 (2018).

20. Latysheva, N. S. & Babu, M. M. Discovering and understanding oncogenic gene fusions through data intensive computational approaches. *Nucleic Acids Res* 44, 4487–4503 (2016).

21. Kolomietz, E., Meyn, M. S., Pandita, A. & Squire, J. A. The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumors. *Genes, Chromosomes and Cancer* 35, 97–112 (2002).

22. Shugay, M., Ortiz de Mendíbil, I., Vizmanos, J. L. & Novo, F. J. Genomic Hallmarks of Genes Involved in Chromosomal Translocations in Hematological Cancer. *PLOS Computational Biology* 8, e1002797 (2012).

23. Zhao, Y. *et al.* Distinct distributions of genomic features of the 5' and 3' partners of coding somatic cancer gene fusions: arising mechanisms and functional implications. *Oncotarget* 8, 66769–66783 (2017).

24. Liu, S. *et al.* Comprehensive evaluation of fusion transcript detection algorithms and a meta-caller to combine top performing methods in paired-end RNA-seq data. *Nucleic Acids Research* 44, e47–e47 (2016).

25. Wang, Q., Xia, J., Jia, P., Pao, W. & Zhao, Z. Application of next generation sequencing to human gene fusion detection: computational tools, features and perspectives. *Briefings in Bioinformatics* 14, 506–519 (2013).

26. Schadt, E. E., Turner, S. & Kasarskis, A. A window into third-generation sequencing. *Human Molecular Genetics* 19, R227–R240 (2010).

27. Rhoads, A. & Au, K. F. PacBio Sequencing and its Applications. *Genomics, Proteomics & Bioinformatics* 13, 278–289 (2015).

28. Eid, J. *et al.* Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* 323, 133–138 (2009).

29. Wenger, A. M. *et al.* Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology* 37, 1155–1162 (2019).

30. Sondka, Z. *et al.* The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nature Reviews Cancer* 18, 696–705 (2018).

31. Seelig, D. M. *et al.* Canine T-Zone Lymphoma: Unique Immunophenotypic Features, Outcome, and Population Characteristics. *Journal of Veterinary Internal Medicine* 28, 878–886 (2014).

32. Cobbold, S. & Metcalfe, S. Monoclonal antibodies that define canine homologues of human CD antigens: Summary of the First International Canine Leukocyte Antigen Workshop (CLAW). *Tissue Antigens* 43, 137–154 (1994).

33. Abrams, V. K. *et al.* A novel monoclonal antibody specific for canine CD25 (P4A10): Selection and evaluation of canine Tregs. *Veterinary Immunology and Immunopathology* 135, 257–265 (2010).

34. Haas, B. J. *et al.* Targeted in silico characterization of fusion transcripts in tumor and normal tissues via FusionInspector. *Cell Reports Methods* 3, (2023).

35. Wickham, H. *Ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag New York, 2016).

36. R Core Team. *R: A Language and Environment for Statistical Computing*. (R Foundation for Statistical Computing, Vienna, Austria, 2022).

37. Kassambara, A. *Ggpubr: 'ggplot2' Based Publication Ready Plots*. (2023).

38. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology* 16, 284–287 (2012).

39. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* 27, 1739–1740 (2011).

40. Yates, A. D. *et al.* Ensembl 2020. *Nucleic Acids Research* 48, D682–D688 (2020).

41. Cunningham, F. *et al.* Ensembl 2022. *Nucleic Acids Research* 50, D988–D995 (2022).

42. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34, 3094–3100 (2018).

43. Perez, G. *et al.* The UCSC Genome Browser database: 2025 update. *Nucleic Acids Research* 53, D1243–D1249 (2025).

44. Diesh, C. *et al.* JBrowse 2: a modular genome browser with views of synteny and structural variation. *Genome Biology* 24, 74 (2023).

45. Pedersen, L. Ø. *et al.* The interaction of beta 2-microglobulin (β2m) with mouse class I major histocompatibility antigens and its ability to support peptide binding. A comparison of human and mouse β2m. *European Journal of Immunology* 25, 1609–1616 (1995).

46. Sebbag-Sznajder, N. *et al.* Regulation of alternative splicing within the supraspliceosome. *Journal of Structural Biology* 177, 152–159 (2012).

47. Yang, S., Jia, R. & Bian, Z. SRSF5 functions as a novel oncogenic splicing factor and is upregulated by oncogene SRSF3 in oral squamous cell carcinoma. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* 1865, 1161–1172 (2018).

48. Paul, A. M. *et al.* The Revelation of Continuously Organized, Co-Overexpressed Protein-Coding Genes with Roles in Cellular Communications in Breast Cancer. *Cells* 11, (2022).

49. Jin, D.-H. *et al.* Overexpression of MAPK15 in gastric cancer is associated with copy number gain and contributes to the stability of c-Jun. *Oncotarget; Vol 6, No 24* (2015).

50. Mei, L. & Nave, K.-A. Neuregulin-ERBB Signaling in the Nervous System and Neuropsychiatric Diseases. *Neuron* 83, 27–49 (2014).

51. Li, A. *et al.* TOX2 nuclear-cytosol translocation is linked to leukemogenesis of acute T-cell leukemia by repressing TIM3 transcription. *Cell Death & Differentiation* 31, 1506–1518 (2024).

52. Matthews, J. M., Lester, K., Joseph, S. & Curtis, D. J. LIM-domain-only proteins in cancer. *Nature Reviews Cancer* 13, 111–122 (2013).

53. Wang, N., Dong, Q. & Zhou, X.-N. LMO4 promotes the invasion and proliferation of gastric cancer by activating PI3K-Akt-mTOR signaling. *Am J Transl Res* 11, 6534–6543 (2019).

54. Wang, W., Wu, S., Guo, M. & He, J. LMO4 is a prognostic marker involved in cell migration and invasion in non-small-cell lung cancer. *Journal of Thoracic Disease; Vol 8, No 12 (December 31, 2016): Journal of Thoracic Disease* (2016).

55. Zhang, Y. *et al.* Krüppel-like Factor 6 Is a Co-activator of NF-κB That Mediates p65-dependent Transcription of Selected Downstream Genes *. *Journal of Biological Chemistry* 289, 12876–12885 (2014).

56. Wu, S., Zhou, B., Xu, L. & Sun, H. Interaction between nuclear insulin receptor substrate-2 and NF-κB in IGF-1 induces response in breast cancer cells. *Oncol Rep* 24, 1541–1550 (2010).

57. Gery, S. *et al.* The Circadian Gene Per1 Plays an Important Role in Cell Growth and DNA Damage Control in Human Cancer Cells. *Molecular Cell* 22, 375–382 (2006).

58. Coni, S. *et al.* Blockade of EIF5A hypusination limits colorectal cancer growth by inhibiting MYC elongation. *Cell Death & Disease* 11, 1045 (2020).

59. Fujimura, K. *et al.* Eukaryotic Translation Initiation Factor 5A (EIF5A) Regulates Pancreatic Cancer Metastasis by Modulating RhoA and Rho-associated Kinase (ROCK) Protein Expression Levels *. *Journal of Biological Chemistry* 290, 29907–29919 (2015).

60. Nyren-Erickson, E. K., Jones, J. M., Srivastava, D. K. & Mallik, S. A disintegrin and metalloproteinase-12 (ADAM12): Function, roles in disease progression, and clinical implications. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1830, 4445–4455 (2013).

61. Kagan, H. M. & Li, W. Lysyl oxidase: Properties, specificity, and biological roles inside and outside of the cell. *Journal of Cellular Biochemistry* 88, 660–672 (2003).

62. Lv, W. *et al.* A novel immune score model predicting the prognosis and immunotherapy response of breast cancer. *Scientific Reports* 13, 6403 (2023).

63. Zhang, D. *et al.* ASCL2 induces an immune excluded microenvironment by activating cancer-associated fibroblasts in microsatellite stable colorectal cancer. *Oncogene* 42, 2841–2853 (2023).

64. Chakravarty, D. *et al.* OncoKB: A Precision Oncology Knowledge Base. *JCO Precis Oncol* 1–16 (2017) doi:10.1200/PO.17.00011.

65. Suehnholz, S. P. *et al.* Quantifying the Expanding Landscape of Clinical Actionability for Patients with Cancer. *Cancer Discovery* 14, 49–65 (2024).

66. Scott, J. S. *et al.* Emerging roles of cohesin-STAG2 in cancer. *Oncogene* 44, 277–287 (2025).

67. Booth, C. *et al.* X-linked lymphoproliferative disease due to SAP/SH2D1A deficiency: a multicenter study on the manifestations, management and outcome of the disease. *Blood* 117, 53–62 (2011).

68. Xiang, Q.-M. *et al.* Overexpression of SH2D1A promotes cancer progression and is associated with immune cell infiltration in hepatocellular carcinoma via bioinformatics and in vitro study. *BMC Cancer* 23, 1005 (2023).

69. Zhang, H. *et al.* PLEKHA5 regulates tumor growth in metastatic melanoma. *Cancer* 126, 1016–1030 (2020).

70. Nagamura, Y. *et al.* PLEKHA5 regulates the survival and peritoneal dissemination of diffuse-type gastric carcinoma cells with Met gene amplification. *Oncogenesis* 10, 25 (2021).

71. Kim, H., Ekram, M. B., Bakshi, A. & Kim, J. AEBP2 as a transcriptional activator and its role in cell migration. *Genomics* 105, 108–115 (2015).

72. Liu, B.-X. *et al.* SERPINB5 promotes colorectal cancer invasion and migration by promoting EMT and angiogenesis via the TNF-α/NF-κB pathway. *International Immunopharmacology* 131, 111759 (2024).

73. Heavican, T. B. *et al.* Genetic drivers of oncogenic pathways in molecular subgroups of peripheral T-cell lymphoma. *Blood* 133, 1664–1676 (2019).

74. Herek, T. A. *et al.* DNMT3A mutations define a unique biological and prognostic subgroup associated with cytotoxic T cells in PTCL-NOS. *Blood* 140, 1278–1290 (2022).

75. Schatz, J. H. *et al.* Targeted mutational profiling of peripheral T-cell lymphoma not otherwise specified highlights new mechanisms in a heterogeneous pathogenesis. *Leukemia* 29, 237–241 (2015).

76. Pan, B. *et al.* Similarities and differences between variants called with human reference genome HG19 or HG38. *BMC Bioinformatics* 20, 101 (2019).

77. Wang, C. *et al.* A novel canine reference genome resolves genomic architecture and uncovers transcript complexity. *Communications Biology* 4, 185 (2021).

CONCLUSIONS

The dismal survival times and poor treatment responses reported with peripheral T-cell lymphoma (PTCL) in dogs and humans stress the critical need for improved models to study this aggressive disease. The collective work of this dissertation sought to characterize the gene expression and mutational profile of canine PTCL and investigate its cell of origin to better understand the shared features between the canine and human malignancies. Our findings suggest that canine $CD4^+$ PTCL arises from a thymic precursor cell of origin and is clinically, immunophenotypically, and molecularly similar to human PTCL-NOS, particularly the more aggressive GATA3-PTCL variant.

Similar to human PTCL-NOS, canine $CD4^+$ PTCL exhibited upregulation of *GATA3,* enrichment for PI3K/AKT/mTOR signaling, evidence of reduced PTEN activity, and frequent mutations (in the form of short variants or large structural variants) in *TP53, CDKN2A, PTEN,* and epigenetic modifiers *DNMT3A, TET2,* and *KMT2D*. *GATA3* upregulation, mutations in the *TP53/CDKN2A* axis and *PTEN,* and enrichment for PI3K/AKT/mTOR are particularly characteristic of the GATA3-PTCL subtype that carries a worse prognosis in humans. The results of our study highlight potential candidates for future canine clinical trials that could be translated to human PTCL-NOS. For example, given the frequent mutations in epigenetic modifiers in both species, epigenetic modification-targeting drugs like histone deacetylase (HDAC) inhibitors may be promising candidates, or PI3K inhibitors based on the enrichment of this pathway in both species.

Our study also identified a thymic precursor cell as the most likely cell of origin of canine $CD4^+$ PTCL. Canine $CD4^+$ PTCL was enriched for gene signatures associated with early thymic progenitors, upregulated genes associated with T-cell immaturity, and downregulated TCR signaling, CD25 (*IL2RA*), MHC class II-associated genes, and other markers of T-cell maturity.

These findings have important implications for the pathogenesis of canine PTCL, as it implies that chronic antigen stimulation is likely not a predisposing factor. These findings may also have therapeutic implications, as a thymic precursor cell of origin may indicate dependency of these tumor cells on pathways normally involved in thymocyte survival, like Notch signaling, which could represent another potential therapeutic target. Future studies to validate the *in vitro* and *in vivo* dependency of $CD4^+$ PTCL cells on these pathways for survival and proliferation are necessary to further explore this potential.

As part of our investigation into the cell of origin of $CD4^+$ PTCL, we curated a single-cell transcriptomic atlas of normal canine thymocytes and nodal T cells to capture changes in gene expression as T cells move through different developmental and differentiation stages from the earliest thymic precursors to fully mature, post-activated T cells. This atlas represents a valuable resource that could be used in future studies to more accurately probe the cell of origin of other canine malignancies, such as the more immature T-cell acute lymphocytic leukemia (T-ALL) and the more mature T-zone lymphoma (TZL), without having to rely on publicly available human and murine data that have limitations in their cross-species generalizability.

We were able to validate several fusion gene transcripts predicted from RNA-sequencing data of canine $CD4^+$ PTCL by PCR and Sanger sequencing and determined their prevalence in 120 $CD4^+$ PTCLs. Many of these fusions were also able to be detected in RNA from dogs without lymphoma, emphasizing the importance of rigorous validation of fusion gene candidates predicted from high-throughput sequencing methods. Though our study did not identify fusion genes identical to those that have been described in small subsets of human PTCL-NOS, we did identify one fusion at the RNA level in canine $CD4^+$ PTCL—*TOX2-LMO4*—that showed greater enrichment for NF-κB signaling than $CD4^+$ PTCLs lacking this fusion. The *FYN-TRAF3IP2* fusion gene in PTCL-NOS is known to induce lymphomagenesis through activation of NF-κB signaling. Future studies

characterizing the mechanism driving this enrichment in *TOX2-LMO4*+ $CD4^+$ PTCLs and demonstrating the oncogenic potential of this fusion *in vitro* and *in vivo* may uncover a similar pathogenesis and support the investigation of IκB inhibitors as a novel therapeutic option that may be translatable to human PTCL-NOS.

With long-read DNA sequencing, we were able to identify large chromosomal breaks that corroborated fusion transcripts detected by bulk RNA-sequencing for 3 fusion gene candidates. One of these fusions, *STAG2-SH2D1A*, is a fusion of two distant genes on the X chromosome, and in human males, mutation in *SH2D1A* is associated with the development of X-linked lymphoproliferative disease. Given the overrepresentation of male dogs in $CD4^+$ PTCL, this potential X-linked genetic contribution to tumorigenesis could be an interesting focus for future studies. Another of these fusions was *PLEKHA5-AEBP2*, and the *PLEKHA5* gene's role in mediation of cell signaling has implicated it as both a tumor suppressor gene and an oncogene in various human tumors. Because these fusions were detected in only 1 dog by RNA-sequencing, they were not among the initial candidates for which 120 $CD4^+$ PTCLs were screened by PCR, so our understanding of how recurrent these fusions may be is still unclear. Typically, recurrent fusion genes represent the more likely driver mutations, so screening additional $CD4^+$ PTCLs for these fusions with PCR would be valuable to determine their prevalence and likelihood to be drivers of this disease in dogs. Given the supporting evidence for these fusions at both the DNA and RNA level and involvement of known cancer-associated genes, they are valuable candidates to pursue for future study.

Taken together, our work supports canine $CD4^+$ PTCL as a naturally occurring preclinical model of human PTCL-NOS and challenges the long-held assumption that these tumors arise from a mature nodal T cell. And while we have only scratched the surface of the thousands of fusion gene candidates identified between bulk RNA-sequencing and long-read DNA sequencing data of canine

CD4$^+$ PTCL, our efforts have uncovered a few promising candidates for future study and demonstrated the necessity of multiple molecular assays for the validation of fusion gene candidates. These findings lay the groundwork for future studies of the pathogenesis of CD4$^+$ PTCL and options for targeted therapies that could benefit both the canine and human forms of this malignancy.

APPENDICES

Appendix 1: Supplementary Material for Chapter One

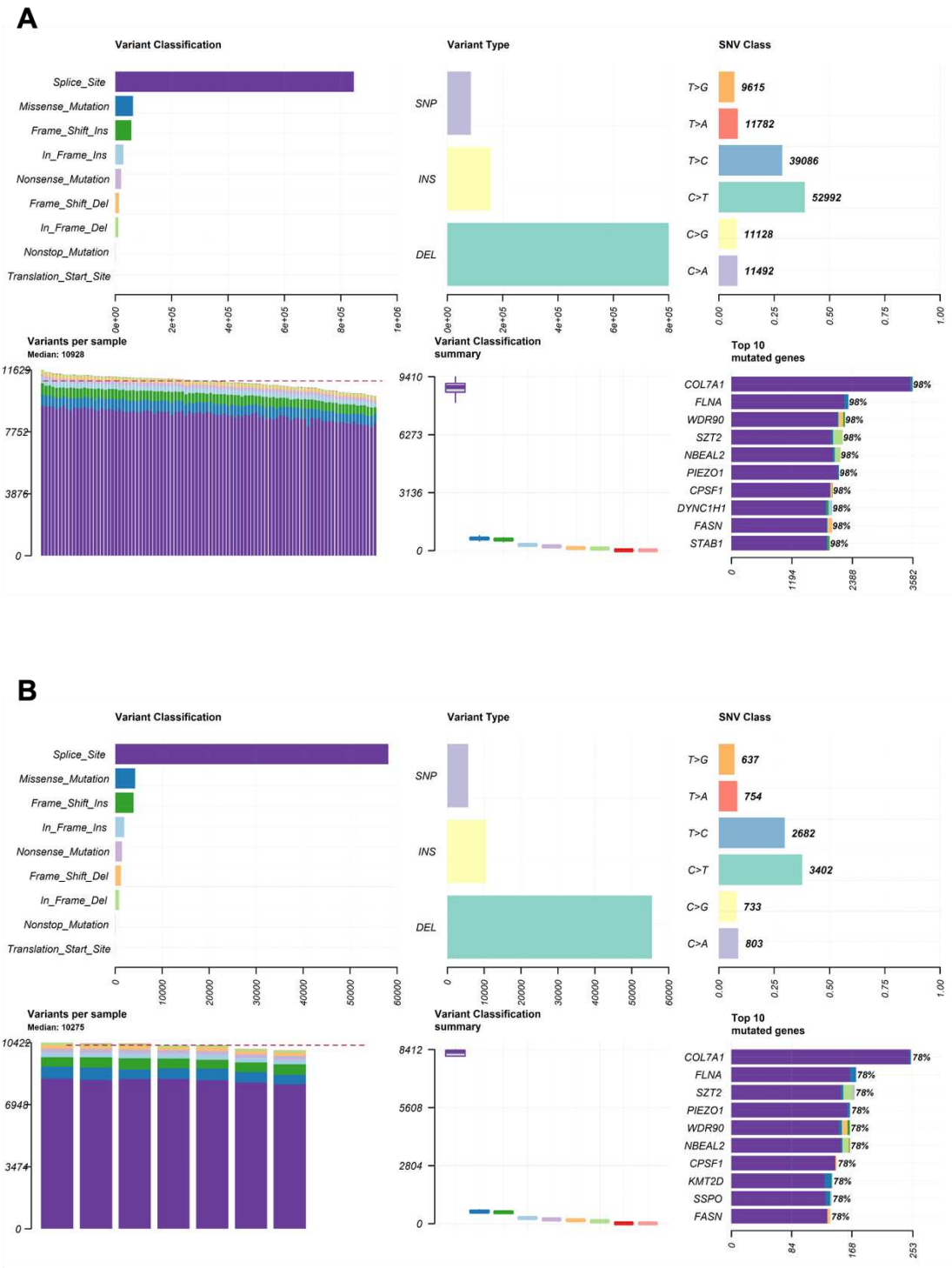
Data availability

Normalized bulk RNA-seq read counts, DESeq2 results for the top 500 differentially expressed genes between canine $CD4^+$ PTCL compared to control $CD4^+$ lymphocytes, and full GSEA and GSVA results for Cohort 1 are available as supplementary data through our publication in *BMC Cancer*: https://doi.org/10.1186/s12885-023-11762-w.

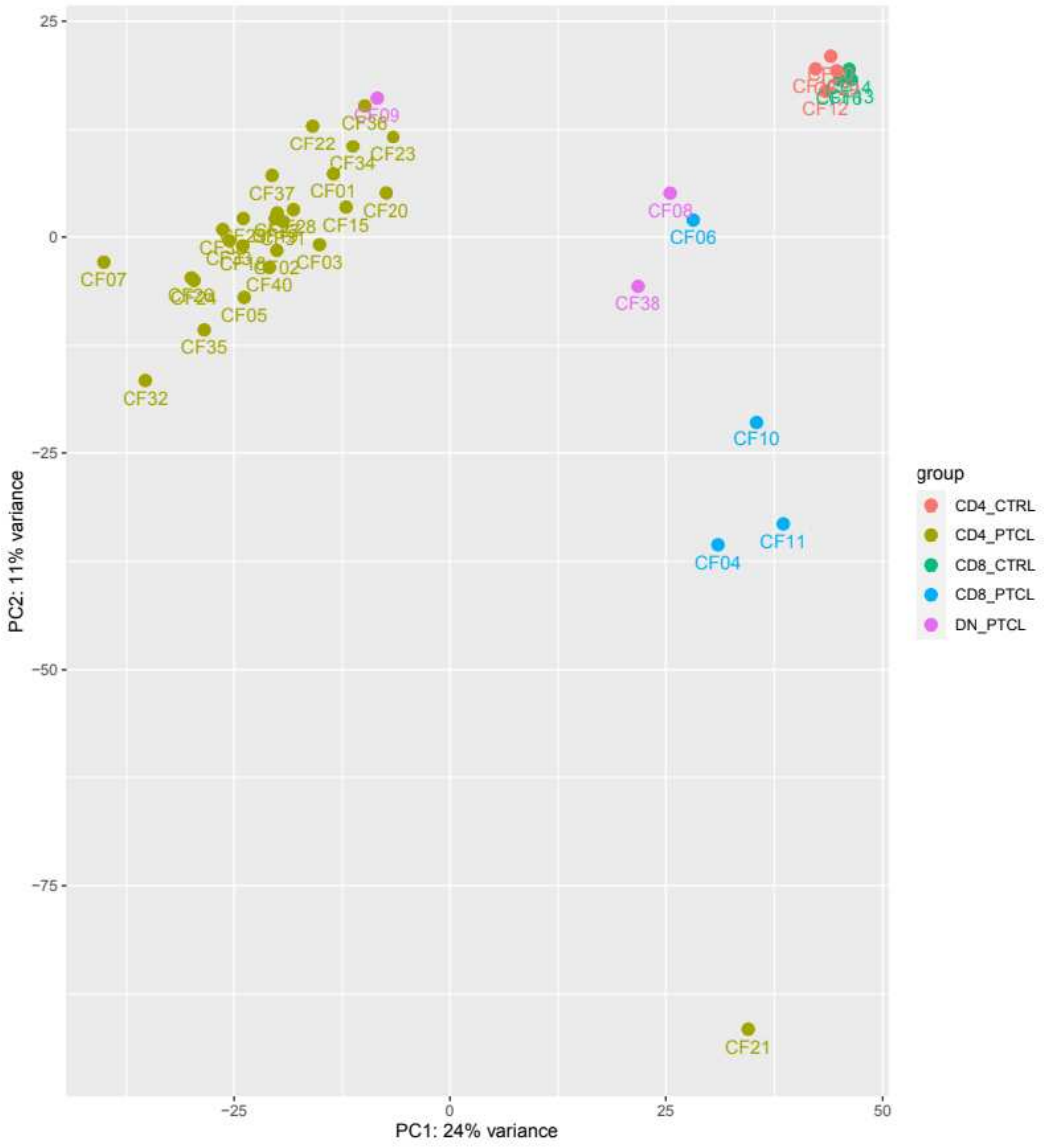
Scripts and software packages used for processing, aligning, and tabulating bulk RNA-seq reads and performing differential expression analysis, hierarchical clustering, and GSEA are documented in a Github repository: https://github.com/pirateyoho/CaninePTCLRNASeq. Scripts and software packages used for short variant calling from bulk RNA-seq data in Cohort 2 are documented in an additional Github repository: https://github.com/pirateyoho/GATKVariantAnalysisRNASeq.

Source code with results for Chapter 1 figures is also available on RPubs:

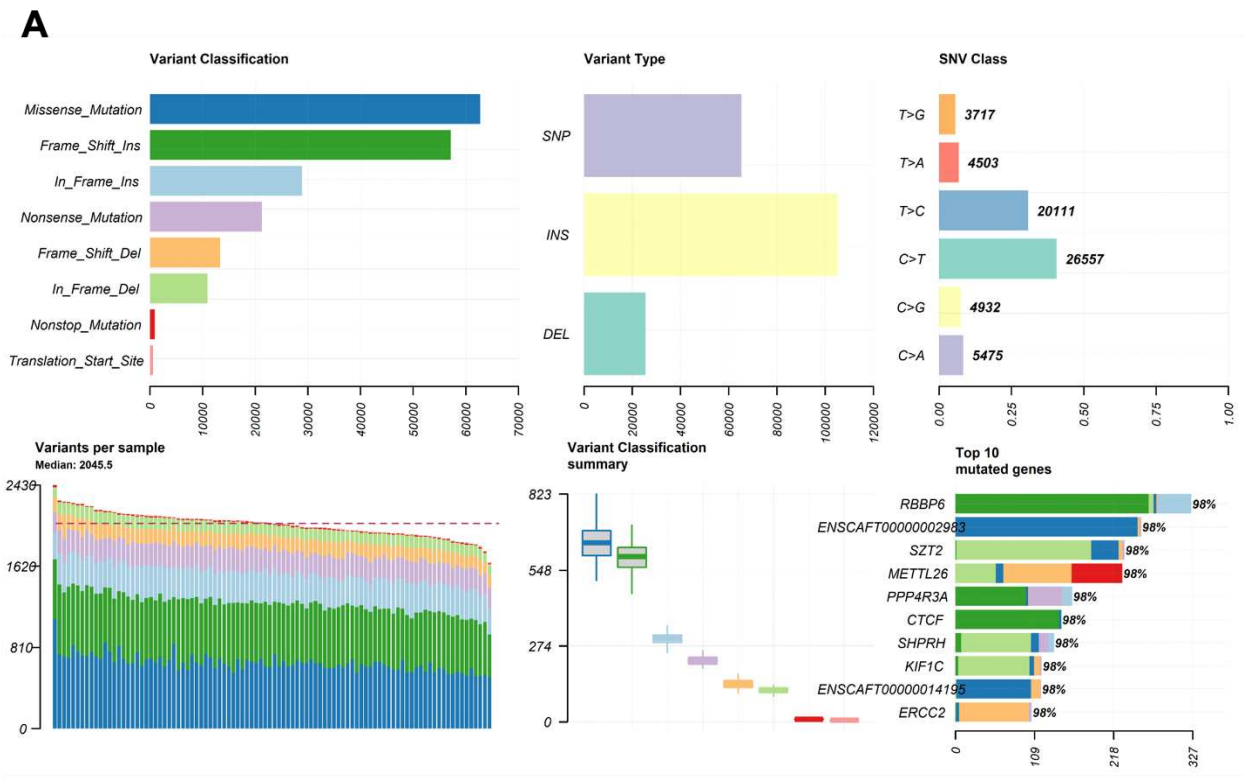
Figure(s)	Source code
Figure 6A, Figure 8A	https://rpubs.com/eileenowens/deseq2_cohort1
Figure 6B, Figure 8B-C, Figure 10E	https://rpubs.com/eileenowens/deseq2_cohort2
Figure 7A-B, Figure 10F-G	https://rpubs.com/eileenowens/caninePTCL_hclust
Figure 9A-B, Figure 11	https://rpubs.com/eileenowens/CD4PTCL_GSEA_GSVA
Figure 12A-E, Figure 13A-B, Figure 14A, Figure 15A, Figure 16A, Supplemental Figure 3	https://rpubs.com/eileenowens/CD4PTCL_maftools
Supplemental Figure 1, **Supplemental Figure 3**	https://rpubs.com/eileenowens/CD4PTCL_maftools_v1
Supplemental Figure 4	https://rpubs.com/eileenowens/hclust-mutations

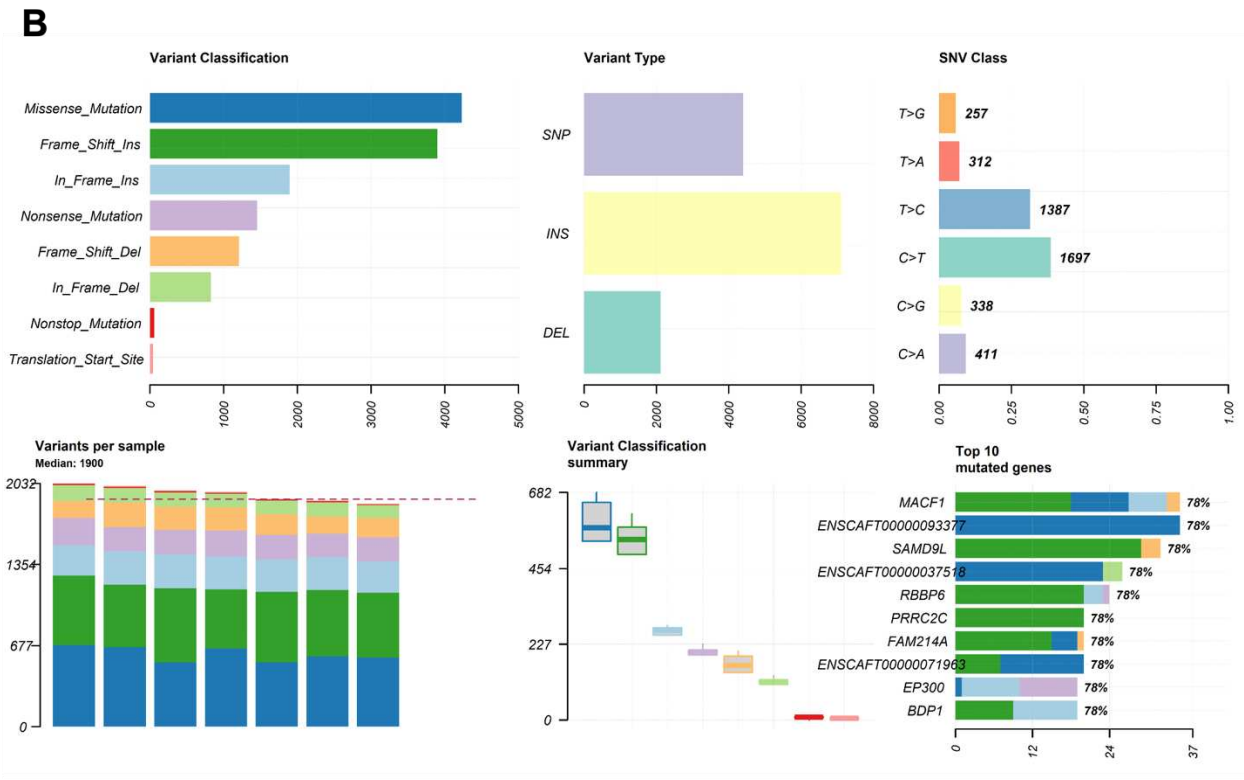


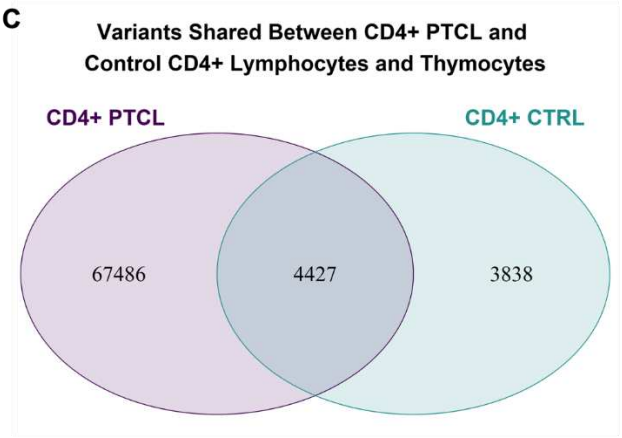
Supplemental Figure 1: The predominance of splice site variant calls (purple) in both CD4⁺ PTCL (**A**) and control CD4⁺ nodal lymphocytes and thymocytes (**B**) following hard filtering with GATK pipeline tools and exclusion of intronic variants and those predicted to have low or modifier effects. These splice site variant calls were suspected to be false positives related to known challenges with calling variants in RNA-seq data and were ultimately filtered out prior to additional downstream analysis.



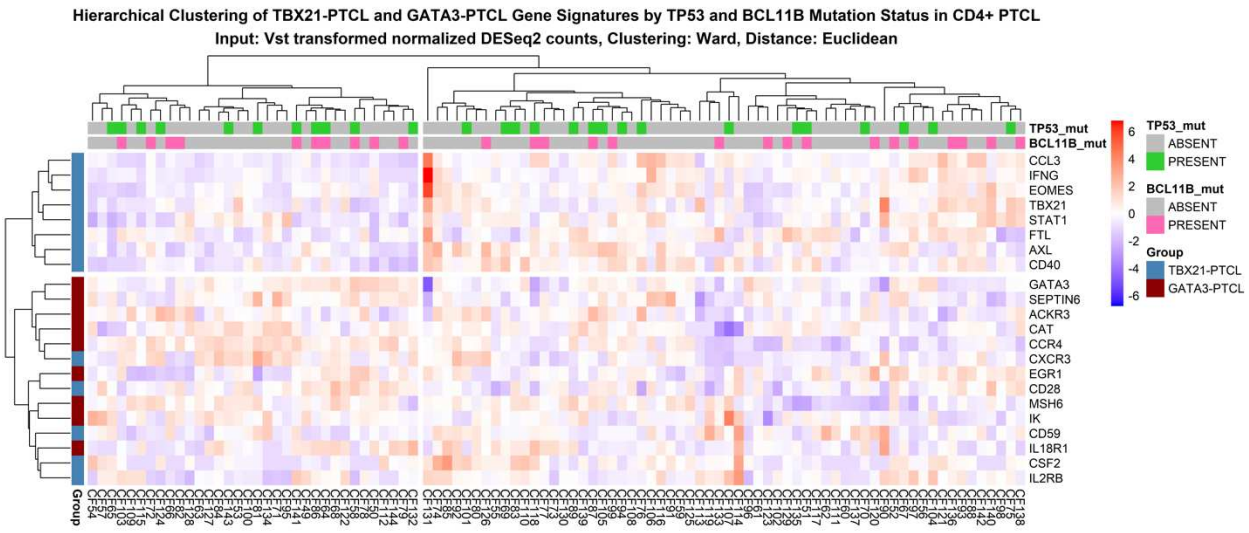
Supplemental Figure 2: Principal component analysis (PCA) of all samples in our study revealed an outlier (CF21) whose variation in the second principal component dimension that was >3x the standard deviation of all samples in the CD4+ PTCL group. This outlier was subsequently removed from further gene expression and gene set enrichment analyses.


A
Variant Classification
Missense_Mutation
Frame_Shift_Ins
In_Frame_Ins
Nonsense_Mutation
Frame_Shift_Del
In_Frame_Del
Nonstop_Mutation
Translation_Start_Site
0
10000
20000
30000
40000
50000
60000
70000
Variant Type
SNP
INS
DEL
0
20000
40000
60000
80000
100000
120000
SNV Class
T>G 3717
T>A 4503
T>C 20111
C>T 26557
C>G 4932
C>A 5475
0.00
0.25
0.50
0.75
1.00
Variants per sample
Median: 2045.5
2430
1620
810
0
Variant Classification summary
823
548
274
0
Top 10 mutated genes
RBBP6 98%
ENSCAFT00000002983 98%
SZT2 98%
METTL26 98%
PPP4R3A 98%
CTCF 98%
SHPRH 98%
KIF1C 98%
ENSCAFT00000014195 98%
ERCC2 98%
0
109
218
327


B
Variant Classification
Missense_Mutation
Frame_Shift_Ins
In_Frame_Ins
Nonsense_Mutation
Frame_Shift_Del
In_Frame_Del
Nonstop_Mutation
Translation_Start_Site
0
1000
2000
3000
4000
5000
Variant Type
SNP
INS
DEL
0
2000
4000
6000
8000
SNV Class
T>G 257
T>A 312
T>C 1387
C>T 1697
C>G 338
C>A 411
0.00
0.25
0.50
0.75
1.00
Variants per sample
Median: 1900
2032
1354
677
0
Variant Classification summary
682
454
227
0
Top 10 mutated genes
MACF1 78%
ENSCAFT00000093377 78%
SAMD9L 78%
ENSCAFT00000037518 78%
RBBP6 78%
PRRC2C 78%
FAM214A 78%
ENSCAFT00000071963 78%
EP300 78%
BDP1 78%
0
12
24
37



Supplemental Figure 3: Summary of small variant calls in CD4⁺ PTCL (**A**) and control CD4⁺ nodal lymphocytes and CD4⁺ thymocytes (**B**) in Cohort 2 following filtering with GATK pipeline tools and excluding variants annotated as intronic, splice site, splice region, and those predicted to have only low or modifier effects, but prior to filtering of any variants shared between tumors and controls. (**C**) 4,427 variants spanning the same start position and end position on the same chromosome and having the same variant classification (nonsense, frameshift, etc.) and variant type (SNP, INDEL) were shared between canine CD4⁺ PTCL samples and control samples. These variants were ultimately filtered out from the CD4⁺ PTCL cohort before proceeding with additional downstream analysis.



Supplemental Figure 4: In human PTCL-NOS, the presence of a *TP53* mutation or gains in *BCL11B* correlate with the GATA3- and TBX21-PTCL molecular subtypes, respectively. In our study, the presence of a *TP53* or *BCL11B* small variant did not correlate with expression of other genes in the GATA3- or TBX21-PTCL signatures.

Supplemental Table 1. Sample information for bulk RNA-seq experiments. F = female, M = male, FS = female spayed, MC = male castrated, LN = lymph node, TH = thymus, MD = mediastinum, BM = bone marrow, LN-MD = mediastinal lymph node, LN-MS = mesenteric lymph node, U = unknown.

Sample name	Experiment group	Breed	Sex	Age (yrs)	Sample location

CF01	Cohort 1 CD4+ PTCL	Miniature Australian Shepherd	MC	5	LN
CF02	Cohort 1 CD4+ PTCL	Shih Tzu	MC	11	LN
CF03	Cohort 1 CD4+ PTCL	Boxer	FS	9	LN
CF04	Cohort 1 CD8+ PTCL	Mixed breed	FS	12	LN
CF05	Cohort 1 CD4+ PTCL	Golden Retriever	FS	10	LN
CF06	Cohort 1 CD8+ PTCL	Mixed breed	MC	5	LN
CF07	Cohort 1 CD4+ PTCL	Mixed breed	MC	11	LN
CF08	Cohort 1 CD4-CD8- PTCL	Cocker Spaniel	FS	3	LN
CF09	Cohort 1 CD4-CD8- PTCL	Mixed breed	FS	4	LN
CF10	Cohort 1 CD8+ PTCL	Mixed breed	FS	13	LN
CF11	Cohort 1 CD8+ PTCL	Lhasa Apso	MC	U	LN
CF12	Cohort 1 Control CD4+ nodal lymphocytes	Beagle, Beagle	M, F	1, 1	LN
CF13	Cohort 1 Control CD8+ nodal lymphocytes	Hound, Hound	F, F	1, 1	LN
CF14	Cohort 1 Control CD8+ nodal lymphocytes	Hound, Hound	F, F	1, 1	LN
CF15	Cohort 1 CD4+ PTCL	Mixed breed	MC	6	LN
CF16	Cohort 1 Control CD8+ nodal lymphocytes	Hound	F	1	LN
CF17	Cohort 1 Control CD4+ nodal lymphocytes	Hound	F	1	LN
CF18	Cohort 1 CD4+ PTCL	Boxer	MC	5	LN
CF19	Cohort 1 CD4+ PTCL	Boxer	FS	4	MD
CF20	Cohort 1 CD4+ PTCL	German Shepherd	FS	6	BM
CF21	Cohort 1 CD4+ PTCL	Labrador Retriever	MC	7	MD
CF22	Cohort 1 CD4+ PTCL	Chesapeake Bay Retriever	MC	7	LN
CF23	Cohort 1 CD4+ PTCL	Golden Retriever	MC	10	LN
CF24	Cohort 1 CD4+ PTCL	Rhodesian Ridgeback	MC	3	MD
CF25	Cohort 1 CD4+ PTCL	Golden Retriever	MC	4	LN
CF26	Cohort 1 CD4+ PTCL	Dogue de Bordeaux	MC	5	LN
CF27	Cohort 1 CD4+ PTCL	Boxer	MC	9	LN
CF28	Cohort 1 CD4+ PTCL	Boxer	FS	10	LN
CF29	Cohort 1 Control CD4+ nodal lymphocytes	Beagle	M	1	LN
CF30	Cohort 1 Control CD4+ nodal lymphocytes	Mixed breed	F	1	LN
CF31	Cohort 1 CD4+ PTCL	Mixed breed	MC	6	LN
CF32	Cohort 1 CD4+ PTCL	Golden Doodle	MC	10	LN
CF33	Cohort 1 CD4+ PTCL	Boxer	MC	7	LN
CF34	Cohort 1 CD4+ PTCL	Golden Retriever	MC	5	LN
CF35	Cohort 1 CD4+ PTCL	Mixed breed	MC	10	LN
CF36	Cohort 1 CD4+ PTCL	Golden Retriever	MC	4	LN

CF37	Cohort 1 CD4+ PTCL	Australian Shepherd	MC	10	LN
CF38	Cohort 1 CD4-CD8- PTCL	Pomeranian	MC	11	LN
CF39	Cohort 1 CD4+ PTCL	Shetland Sheepdog	MC	8	LN
CF40	Cohort 1 CD4+ PTCL	Presa Canario	FS	5	LN
CF41	Thymus Control	Mixed breed	F	1	TH
CF42	Thymus Control	Mixed breed	F	1	TH
CF43	Thymus Control	Mixed breed	F	1	TH
CF44	Thymus Control	Mixed breed	F	1	TH
CF45	Thymus Control	Mixed breed	F	1	TH
CF46	Thymus Control	Mixed breed	F	1	TH
CF47	Thymus Control	Mixed breed	F	2	TH
CF48	Thymus Control	Mixed breed	F	0.58	TH
CF49	Cohort 2 CD4+ PTCL	Mixed breed	MC	8	LN
CF50	Cohort 2 CD4+ PTCL	Papillon	MC	11	MD
CF51	Cohort 2 CD4+ PTCL	Australian Shepherd	FS	5	LN
CF52	Cohort 2 CD4+ PTCL	English Cocker Spaniel	FS	6	LN
CF53	Cohort 2 CD4+ PTCL	Boykin Spaniel	M	6	LN
CF54	Cohort 2 CD4+ PTCL	Labrador retriever	M	4	LN
CF55	Cohort 2 CD4+ PTCL	Mixed breed	MC	2	LN
CF56	Cohort 2 CD4+ PTCL	Yorkshire Terrier	MC	7	LN
CF57	Cohort 2 CD4+ PTCL	Mixed breed	MC	12	LN
CF58	Cohort 2 CD4+ PTCL	Golden Retriever	MC	7	LN
CF59	Cohort 2 CD4+ PTCL	Mixed breed	FS	5	LN
CF60	Cohort 2 CD4+ PTCL	Golden Retriever	MC	5	LN
CF61	Cohort 2 CD4+ PTCL	Mixed breed	MC	4	LN
CF62	Cohort 2 CD4+ PTCL	Boxer	MC	7	LN
CF63	Cohort 2 CD4+ PTCL	Australian Shepherd	MC	5	LN
CF64	Cohort 2 CD4+ PTCL	Saint Bernard	MC	9	LN
CF65	Cohort 2 CD4+ PTCL	Border Collie	MC	7	LN
CF66	Cohort 2 CD4+ PTCL	Boxer	FS	4	LN
CF67	Cohort 2 CD4+ PTCL	Mixed breed	MC	6	LN
CF68	Cohort 2 CD4+ PTCL	Mixed breed	MC	6	LN
CF69	Cohort 2 CD4+ PTCL	Boxer	MC	7	LN
CF70	Cohort 2 CD4+ PTCL	Boxer	F	4	LN
CF71	Cohort 2 CD4+ PTCL	Boxer	FS	7	LN
CF72	Cohort 2 CD4+ PTCL	Old English Sheepdog	FS	4	LN
CF73	Cohort 2 CD4+ PTCL	Labradoodle	MC	5	LN
CF74	Cohort 2 CD4+ PTCL	German Shepherd	MC	7	LN
CF75	Cohort 2 CD4+ PTCL	Mixed breed	FS	6	LN
CF76	Cohort 2 CD4+ PTCL	Pit Bull Terrier	MC	5	LN
CF77	Cohort 2 CD4+ PTCL	Mixed breed	MC	4	LN

CF78	Cohort 2 CD4+ PTCL	Springer Spaniel	MC	6	LN
CF79	Cohort 2 CD4+ PTCL	Siberian Husky	MC	4	LN
CF80	Cohort 2 CD4+ PTCL	Australian Shepherd	MC	9	LN
CF81	Cohort 2 CD4+ PTCL	Mixed breed	FS	5	LN-MS
CF82	Cohort 2 CD4+ PTCL	Bullmastiff	F	4	LN
CF83	Cohort 2 CD4+ PTCL	Boxer	MC	4	LN
CF84	Cohort 2 CD4+ PTCL	Mixed breed	MC	6	LN
CF85	Cohort 2 CD4+ PTCL	Welsh Corgi	FS	8	LN
CF86	Cohort 2 CD4+ PTCL	Golden Doodle	FS	6	LN
CF87	Cohort 2 CD4+ PTCL	American Staffordshire terrier	MC	3	LN
CF88	Cohort 2 CD4+ PTCL	Boxer	FS	5	LN
CF89	Cohort 2 CD4+ PTCL	Mixed breed	MC	8	LN
CF90	Cohort 2 CD4+ PTCL	Shih Tzu	MC	9	LN
CF91	Cohort 2 CD4+ PTCL	English Setter	F	5	LN
CF92	Cohort 2 CD4+ PTCL	Mixed breed	MC	6	LN
CF93	Cohort 2 CD4+ PTCL	Boxer	FS	8	LN
CF94	Cohort 2 CD4+ PTCL	Catahoula	MC	7	LN-MS
CF95	Cohort 2 CD4+ PTCL	Golden Retriever	M	5	MD
CF96	Cohort 2 CD4+ PTCL	Airedale Terrier	MC	10	LN
CF97	Cohort 2 CD4+ PTCL	Mixed breed	MC	6	LN
CF98	Cohort 2 CD4+ PTCL	Cane Corso	MC	10	LN
CF99	Cohort 2 CD4+ PTCL	Boxer	MC	7	LN
CF100	Cohort 2 CD4+ PTCL	Mixed breed	MC	6	LN
CF101	Cohort 2 CD4+ PTCL	Labrador retriever	FS	10	LN
CF102	Cohort 2 CD4+ PTCL	Golden Doodle	MC	4	LN
CF103	Cohort 2 CD4+ PTCL	Boxer	FS	10	LN
CF104	Cohort 2 CD4+ PTCL	Weimaraner	FS	5	LN
CF105	Cohort 2 CD4+ PTCL	Boxer	MC	5	LN
CF106	Cohort 2 CD4+ PTCL	German Shepherd	FS	10	LN
CF107	Cohort 2 CD4+ PTCL	Mixed breed	MC	7	LN
CF108	Cohort 2 CD4+ PTCL	Boxer	MC	10	LN
CF109	Cohort 2 CD4+ PTCL	Australian Cattle Dog	MC	4	LN
CF110	Cohort 2 CD4+ PTCL	Boxer	FS	5	LN
CF111	Cohort 2 CD4+ PTCL	Yorkshire Terrier	FS	9	LN
CF112	Cohort 2 CD4+ PTCL	Welsh Corgi	MC	5	LN-MD
CF113	Cohort 2 CD4+ PTCL	Boxer	FS	4	LN
CF114	Cohort 2 CD4+ PTCL	Mixed breed	FS	11	LN
CF115	Cohort 2 CD4+ PTCL	Australian Shepherd	FS	5	LN
CF116	Cohort 2 CD4+ PTCL	Mixed breed	FS	5	LN
CF117	Cohort 2 CD4+ PTCL	Mixed breed	MC	8	LN
CF118	Cohort 2 CD4+ PTCL	Mixed breed	FS	3	LN

CF119	Cohort 2 CD4+ PTCL	Boxer	MC	8	LN
CF120	Cohort 2 CD4+ PTCL	Boxer	FS	7	LN
CF121	Cohort 2 CD4+ PTCL	Boxer	FS	5	LN
CF122	Cohort 2 CD4+ PTCL	Golden Retriever	FS	6	LN
CF123	Cohort 2 CD4+ PTCL	Mixed breed	MC	8	LN
CF124	Cohort 2 CD4+ PTCL	Blue Heeler	FS	4	LN
CF125	Cohort 2 CD4+ PTCL	Pit Bull Terrier	FS	3	LN
CF126	Cohort 2 CD4+ PTCL	American Bulldog	FS	5	LN
CF127	Cohort 2 CD4+ PTCL	Golden Retriever	MC	8	LN
CF128	Cohort 2 CD4+ PTCL	Boxer	MC	7	LN
CF129	Cohort 2 CD4+ PTCL	Boxer	MC	7	LN
CF130	Cohort 2 CD4+ PTCL	Australian Shepherd	MC	12	LN
CF131	Cohort 2 CD4+ PTCL	Mixed breed	MC	13	LN
CF132	Cohort 2 CD4+ PTCL	Mixed breed	MC	8	LN
CF133	Cohort 2 CD4+ PTCL	Boxer	FS	6	LN
CF134	Cohort 2 CD4+ PTCL	Old English Sheepdog	MC	3	LN
CF135	Cohort 2 CD4+ PTCL	Mixed breed	MC	4	LN
CF136	Cohort 2 CD4+ PTCL	Mixed breed	MC	5	LN
CF137	Cohort 2 CD4+ PTCL	Boxer	MC	7	LN
CF138	Cohort 2 CD4+ PTCL	Golden Doodle	FS	8	LN
CF139	Cohort 2 CD4+ PTCL	Mixed breed	FS	6	LN
CF140	Cohort 2 CD4+ PTCL	Boxer	MC	9	LN
CF141	Cohort 2 CD4+ PTCL	Boxer	MC	11	LN
CF142	Cohort 2 CD4+ PTCL	Newfoundland	MC	4	LN
CF143	Cohort 2 CD4+ PTCL	Mixed breed	M-	5	LN
CF144	Cohort 2 CD4+ PTCL	Golden Doodle	MC	7	LN
CF145	Cohort 2 Control CD4+ nodal lymphocytes	Mixed breed	F	1	LN
CF146	Cohort 2 Control CD4+ nodal lymphocytes	Mixed breed	F	1	LN
CF147	Cohort 2 Control CD4+ nodal lymphocytes	Mixed breed	F	1	LN
CF148	Cohort 2 Control CD4+ thymocytes	Mixed breed	F	0.66	TH
CF149	Cohort 2 Control CD4+ nodal lymphocytes	Mixed breed	F	0.66	LN
CF150	Cohort 2 Control CD4+ nodal lymphocytes	Mixed breed	F	0.73	LN
CF151	Cohort 2 Control CD4+ thymocytes	Mixed breed	F	0.73	TH

Supplemental Table 2: Flow cytometric features of canine CD4+ PTCL.

Sample name	Experiment group	Viability (%)	% Neoplastic	% CD25+	MHC CL II MFI	CD5 MFI	CD3 MFI	CD4 MFI
CF01	Cohort 1 CD4+ PTCL	70.95	89.42	0.15	3.71	18.1	4.86	5.45
CF02	Cohort 1 CD4+ PTCL	82.15	84.31	0.08	0.42	3.08	11.84	3.98
CF03	Cohort 1 CD4+ PTCL	91.85	93.68	0.11	1.03	30.46	13.17	4.62
CF05	Cohort 1 CD4+ PTCL	92.78	95.73	0.01	1.25	0.25	35.59	5.78
CF07	Cohort 1 CD4+ PTCL	87.87	94.83	0.03	3.58	51.44	25.05	4.4
CF15	Cohort 1 CD4+ PTCL	93.7	89.3	0.31	6.8	87.54	10.72	4.54
CF18	Cohort 1 CD4+ PTCL	76.02	84.24	0.13	14.16	0.29	14.28	7.2
CF19	Cohort 1 CD4+ PTCL	92.23	92.32	0.03	0.74	54.62	14.34	8.16
CF20	Cohort 1 CD4+ PTCL	92.35	85.06	0.08	0.33	20.55	8.66	8.57
CF21	Cohort 1 CD4+ PTCL	69.9	87.74	15.51	1.2	1.68	15.57	14.52
CF22	Cohort 1 CD4+ PTCL	93.95	77.64	0.07	0.32	24.79	5.84	7.83
CF23	Cohort 1 CD4+ PTCL	98.58	94	0.72	0.47	0.25	1.43	3.73
CF24	Cohort 1 CD4+ PTCL	86.68	75.12	0.58	0.4	25.19	4.9	3.79
CF25	Cohort 1 CD4+ PTCL	94.53	94.97	0.06	0.66	0.3	1.45	6.83
CF26	Cohort 1 CD4+ PTCL	94.88	72.31	0.49	0.71	2.11	10.8	4.52
CF27	Cohort 1 CD4+ PTCL	91.79	92.85	0.15	0.85	23.14	14.08	8.81
CF28	Cohort 1 CD4+ PTCL	79.85	92.38	0.26	3.88	61.26	15.25	9.49
CF31	Cohort 1 CD4+ PTCL	89.88	86	0.01	0.38	0.72	9.13	4.57
CF32	Cohort 1 CD4+ PTCL	73.32	66.41	0.15	4.5	26.35	11.96	1.99
CF33	Cohort 1 CD4+ PTCL	94.23	90.72	0.06	0.71	0.33	14.92	6.68
CF34	Cohort 1 CD4+ PTCL	86.73	91.54	0	0.38	0.26	0.73	5.75
CF35	Cohort 1 CD4+ PTCL	88.1	95.05	0.12	6.97	35.18	18.7	3.56
CF36	Cohort 1 CD4+ PTCL	94.9	97.71	0.06	0.38	79	11.11	5.63
CF37	Cohort 1 CD4+ PTCL	96.5	96.79	0.12	0.23	0.29	17.53	0.71
CF39	Cohort 1 CD4+ PTCL	90.55	92.79	0.48	0.59	35.62	13.71	3.14
CF40	Cohort 1 CD4+ PTCL	89.2	96.65	0.03	1.04	50.47	12.34	3.52
CF49	Cohort 2 CD4+ PTCL	92.68	81.92	0.01	0.61	36.25	10.83	4.48
CF50	Cohort 2 CD4+ PTCL	94.55	80.94	0.02	0.21	0.24	8.79	2.33
CF51	Cohort 2 CD4+ PTCL	91.78	94.55	0.06	0.94	48.76	29.63	5.64
CF52	Cohort 2 CD4+ PTCL	92.48	95.26	0.15	0.53	68.97	11.64	3.76
CF53	Cohort 2 CD4+ PTCL	89.85	97.77	0.38	0.53	0.61	14.95	7
CF54	Cohort 2 CD4+ PTCL	95.78	96.15	0	0.34	0.3	15.94	4.73
CF55	Cohort 2 CD4+ PTCL	76.64	89.31	0.02	0.54	0.22	16.13	5.99
CF56	Cohort 2 CD4+ PTCL	87.47	82.93	0.04	3.85	0.31	10.01	3.21
CF57	Cohort 2 CD4+ PTCL	73.57	97.94	0.03	0.73	0.18	18.54	4.12
CF58	Cohort 2 CD4+ PTCL	82.66	98.25	0	0.24	0.21	10.46	3.03
CF59	Cohort 2 CD4+ PTCL	88.64	81.34	0.21	0.67	37.25	11.91	2.51
CF60	Cohort 2 CD4+ PTCL	96.85	94.46	0.03	0.76	56.39	17.2	5.27

CF61	Cohort 2 CD4+ PTCL	91.87	98.26	0.03	0.42	0.42	20.18	5.88
CF62	Cohort 2 CD4+ PTCL	81.21	95.39	0.02	0.55	0.58	18.71	8.57
CF63	Cohort 2 CD4+ PTCL	81.39	98.21	0.01	5.32	67.8	12.22	4.71
CF64	Cohort 2 CD4+ PTCL	70.47	88.3	0.03	0.99	0.46	26.76	3.1
CF65	Cohort 2 CD4+ PTCL	95.59	83.34	0.13	0.49	0.29	19.39	5.92
CF66	Cohort 2 CD4+ PTCL	89.4	95.02	0.09	8.52	22.74	12.71	5.79
CF67	Cohort 2 CD4+ PTCL	90.21	78.59	N/A	1.95	0.85	10.52	6.28
CF68	Cohort 2 CD4+ PTCL	95.99	92.26	N/A	0.27	0.86	11.43	7.12
CF69	Cohort 2 CD4+ PTCL	80.44	77.98	N/A	4.98	9.96	9.26	2.73
CF70	Cohort 2 CD4+ PTCL	88.56	82.1	N/A	1.44	15.28	14.22	5.7
CF71	Cohort 2 CD4+ PTCL	95.27	73.45	N/A	0.27	0.62	9.76	4.59
CF72	Cohort 2 CD4+ PTCL	36.95	74.61	N/A	1.56	0.88	9.54	10.92
CF73	Cohort 2 CD4+ PTCL	85.34	87.56	N/A	1.38	0.43	20.65	4.25
CF74	Cohort 2 CD4+ PTCL	87.72	71.54	N/A	14.15	30.61	7.65	7.51
CF75	Cohort 2 CD4+ PTCL	89.53	63.71	N/A	7.08	11.72	8.43	4.34
CF76	Cohort 2 CD4+ PTCL	40.49	73.24	N/A	2.85	10.47	10.51	4.69
CF77	Cohort 2 CD4+ PTCL	90.04	88.03	N/A	4.87	17.25	5.95	8.35
CF78	Cohort 2 CD4+ PTCL	89.34	45.26	N/A	0.78	0.35	12.57	3.41
CF79	Cohort 2 CD4+ PTCL	93.49	29.42	N/A	0.22	0.31	13.63	2.33
CF80	Cohort 2 CD4+ PTCL	91.2	85.73	N/A	14.03	11.28	7.83	11.4
CF81	Cohort 2 CD4+ PTCL	90.86	90.2	N/A	5.67	22.92	6.65	10.4
CF82	Cohort 2 CD4+ PTCL	95.48	93.81	N/A	3.83	15.23	11.11	9.5
CF83	Cohort 2 CD4+ PTCL	92.26	90.06	N/A	0.93	0.53	14.79	4.55
CF84	Cohort 2 CD4+ PTCL	91.78	93.48	N/A	6.78	21.17	7.13	5.61
CF85	Cohort 2 CD4+ PTCL	94.6	91.22	N/A	2.69	14.76	14.05	6.21
CF86	Cohort 2 CD4+ PTCL	95.52	82.88	N/A	0.33	0.37	17.89	3.37
CF87	Cohort 2 CD4+ PTCL	96.61	97.19	N/A	2.99	12.22	12.4	6.75
CF88	Cohort 2 CD4+ PTCL	86.15	92.6	N/A	7.89	18.07	11.04	4.85
CF89	Cohort 2 CD4+ PTCL	93.51	88.43	N/A	3.5	15.21	10.22	7.63
CF90	Cohort 2 CD4+ PTCL	94.79	83.8	N/A	23.02	12.1	4.51	6.46
CF91	Cohort 2 CD4+ PTCL	92.56	92.84	N/A	2.11	12.27	6.79	6.77
CF92	Cohort 2 CD4+ PTCL	93.12	73.32	N/A	0.57	2.7	11.54	4.72
CF93	Cohort 2 CD4+ PTCL	97.09	94.39	N/A	0.65	0.66	23.77	7.59
CF94	Cohort 2 CD4+ PTCL	95.68	79.93	N/A	0.93	0.76	9.42	10.03
CF95	Cohort 2 CD4+ PTCL	95.13	94.1	N/A	2.34	0.67	13.11	8.1
CF96	Cohort 2 CD4+ PTCL	95.75	96.46	N/A	3.13	15.23	16.38	5.51
CF97	Cohort 2 CD4+ PTCL	94.3	98.05	N/A	6.75	0.64	18.23	4.89
CF98	Cohort 2 CD4+ PTCL	85	95.21	N/A	3.65	10.59	7.51	3.78
CF99	Cohort 2 CD4+ PTCL	81.35	83.87	N/A	11.4	25.08	15.18	4.45
CF100	Cohort 2 CD4+ PTCL	92.81	91.6	N/A	1.15	1.22	14.13	5
CF101	Cohort 2 CD4+ PTCL	93.9	84	N/A	4.45	0.64	9.53	5.97

CF102	Cohort 2 CD4+ PTCL	98.98	81.54	N/A	2.68	0.59	20.25	5.18
CF103	Cohort 2 CD4+ PTCL	93.32	98.12	N/A	2.8	13.63	14.36	6.95
CF104	Cohort 2 CD4+ PTCL	96.07	73.24	N/A	6.45	10.63	18.27	8.13
CF105	Cohort 2 CD4+ PTCL	92.09	83.51	N/A	0.36	18.45	15.91	6.02
CF106	Cohort 2 CD4+ PTCL	87.35	91.04	N/A	4.81	0.7	15.41	8.95
CF107	Cohort 2 CD4+ PTCL	42.22	53.54	N/A	6.4	1.14	6.51	7.29
CF108	Cohort 2 CD4+ PTCL	91.16	50.32	N/A	2.48	0.37	17.97	1.67
CF109	Cohort 2 CD4+ PTCL	86.34	98.18	N/A	0.1	32.38	7	10.52
CF110	Cohort 2 CD4+ PTCL	96.67	63.64	N/A	0.68	35	15.02	5.58
CF111	Cohort 2 CD4+ PTCL	94.74	92.99	N/A	2.47	0.65	13.77	6.05
CF112	Cohort 2 CD4+ PTCL	93.47	96.15	N/A	0.26	25.49	23.25	6.61
CF113	Cohort 2 CD4+ PTCL	91.8	94.09	N/A	0.69	22.8	20.17	9.07
CF114	Cohort 2 CD4+ PTCL	67.94	87.07	N/A	30.41	13.42	0.54	10.82
CF115	Cohort 2 CD4+ PTCL	89.35	97.26	N/A	1.35	12.31	23.84	13.21
CF116	Cohort 2 CD4+ PTCL	86.1	88.58	N/A	1.06	13.38	22.68	2.62
CF117	Cohort 2 CD4+ PTCL	87.34	70.64	N/A	2.91	35.45	12.57	7.78
CF118	Cohort 2 CD4+ PTCL	93.81	84.18	N/A	1.13	34.85	11.54	11.72
CF119	Cohort 2 CD4+ PTCL	90.83	77.72	N/A	1.38	3.31	12.85	4.09
CF120	Cohort 2 CD4+ PTCL	76.66	96.57	N/A	4	29.02	18.85	5.11
CF121	Cohort 2 CD4+ PTCL	76.16	96.19	N/A	2.49	25.77	15.46	4.1
CF122	Cohort 2 CD4+ PTCL	97.58	74.09	N/A	0.37	0.48	12.36	4.9
CF123	Cohort 2 CD4+ PTCL	91.01	96.25	N/A	0.22	25.59	24.79	5.35
CF124	Cohort 2 CD4+ PTCL	93.71	79.47	N/A	0.65	0.32	13.3	6.4
CF125	Cohort 2 CD4+ PTCL	83.75	74.78	N/A	2.05	1.45	39.39	20.44
CF126	Cohort 2 CD4+ PTCL	86.04	74.64	N/A	1.53	22.45	6.59	6.26
CF127	Cohort 2 CD4+ PTCL	82.51	97.95	N/A	1.35	0.84	16.46	7.64
CF128	Cohort 2 CD4+ PTCL	42.08	93.1	N/A	0.68	20.71	8.55	3.6
CF129	Cohort 2 CD4+ PTCL	94.23	95.94	N/A	1.09	18.13	18.2	8.59
CF130	Cohort 2 CD4+ PTCL	76.3	69.02	N/A	1.09	0.49	10.11	3.61
CF131	Cohort 2 CD4+ PTCL	84.08	89.03	N/A	3.85	0.85	7.31	11.65
CF132	Cohort 2 CD4+ PTCL	97.11	26.5	N/A	1.13	15.25	14.83	6.57
CF133	Cohort 2 CD4+ PTCL	90.46	67.41	N/A	6.14	27.01	10.64	7.14
CF134	Cohort 2 CD4+ PTCL	85.66	95.66	N/A	1.08	20.99	9.44	9.14
CF135	Cohort 2 CD4+ PTCL	64.25	89.49	N/A	1.37	6.83	8.96	4.9
CF136	Cohort 2 CD4+ PTCL	86.61	78.06	N/A	0.99	14.36	9.14	6.75
CF137	Cohort 2 CD4+ PTCL	94.45	96.26	N/A	1.37	12.01	20.08	8.05
CF138	Cohort 2 CD4+ PTCL	92.66	97.26	N/A	10.5	31.59	9.78	7.46
CF139	Cohort 2 CD4+ PTCL	67.52	94.02	N/A	5.47	0.66	17.99	8.95
CF140	Cohort 2 CD4+ PTCL	65.81	97.09	N/A	3.71	38.96	11.54	6.9
CF141	Cohort 2 CD4+ PTCL	93.04	88.82	N/A	1.96	38.62	19	5.64
CF142	Cohort 2 CD4+ PTCL	80.5	88.24	N/A	2.98	21.98	12.57	5.33

CF143	Cohort 2 CD4+ PTCL	97.11	94.69	N/A	1.63	12.15	14.17	8.94
CF144	Cohort 2 CD4+ PTCL	94.91	73.47	N/A	0.84	23.06	16.95	4.38

Supplemental Table 3: Novogene RNA-seq QA/QC report.

ID	Raw reads	Clean reads	Raw bases	Clean bases	Error rate (%)	Q20 (%)	Q30 (%)	GC content (%)
CF01	49548966	47959446	7.4G	7.2G	0.03	97.68	93.62	50.23
CF02	48707294	47190500	7.3G	7.1G	0.03	97.52	93.31	50.58
CF03	52926966	51138812	7.9G	7.7G	0.03	97.73	93.85	49.89
CF04	61848274	59937988	9.3G	9G	0.03	97.64	93.69	51.12
CF05	64708040	63022640	9.7G	9.5G	0.03	97.7	93.8	49.92
CF06	54531732	52901242	8.2G	7.9G	0.03	97.86	94.1	50.54
CF07	54004842	52430952	8.1G	7.9G	0.03	97.76	93.97	50.53
CF08	57267918	55267200	8.6G	8.3G	0.03	97.64	93.49	49
CF09	63794176	61605864	9.6G	9.2G	0.03	97.7	93.72	50.78
CF10	81854562	79001794	12.3G	11.9G	0.03	97.74	93.83	51.78
CF11	66083172	64224192	9.9G	9.6G	0.03	97.72	93.86	51.77
CF12	71686242	69323150	10.8G	10.4G	0.03	97.79	94.01	51.21
CF13	55899102	54070548	8.4G	8.1G	0.03	97.74	93.88	51.05
CF14	56832496	55029784	8.5G	8.3G	0.03	97.82	94.05	50.77
CF15	46655240	45027580	7G	6.8G	0.03	97.75	93.88	51.37
CF16	51380682	49598884	7.7G	7.4G	0.02	97.94	94.36	51.14
CF17	49892460	48111496	7.5G	7.2G	0.03	97.78	93.97	51.04
CF18	49907106	48214140	7.5G	7.2G	0.03	97.77	93.94	49.84
CF19	56817058	55231380	8.5G	8.3G	0.03	97.91	94.22	50.87
CF20	64997916	62882384	9.7G	9.4G	0.03	97.8	94.03	51.06
CF21	50074724	48315878	7.5G	7.2G	0.03	97.87	94.16	50.61
CF22	62740124	60789720	9.4G	9.1G	0.03	97.78	93.95	51.31
CF23	55758818	53882748	8.4G	8.1G	0.03	97.83	94.03	51.14
CF24	55384306	53636026	8.3G	8G	0.03	97.64	93.65	51.57
CF25	59408924	57461946	8.9G	8.6G	0.03	97.8	94.03	50.51
CF26	60828100	58535460	9.1G	8.8G	0.03	97.25	92.59	50.49
CF27	67319066	65041220	10.1G	9.8G	0.03	97.75	93.88	50.62
CF28	57422874	55688672	8.6G	8.4G	0.03	97.78	93.95	50.82
CF29	68020740	65801830	10.2G	9.9G	0.03	97.92	94.25	51.61
CF30	70986464	68679512	10.6G	10.3G	0.03	97.73	93.81	51.3
CF31	60838638	58502090	9.1G	8.8G	0.03	97.85	94.1	50.9
CF32	67189882	65277158	10.1G	9.8G	0.03	97.61	93.57	51.67
CF33	47484046	46006612	7.1G	6.9G	0.03	97.82	94.14	51.42
CF34	58334394	56604978	8.8G	8.5G	0.03	97.86	94.12	50.4

CF35	66146888	62293732	9.9G	9.3G	0.03	97.78	93.96	50.99
CF36	64768664	62833840	9.7G	9.4G	0.03	97.72	93.86	51.73
CF37	62849258	60771456	9.4G	9.1G	0.03	97.77	93.95	50.46
CF38	76739088	74386306	11.5G	11.2G	0.03	97.86	94.14	51.41
CF39	50640402	49050588	7.6G	7.4G	0.03	97.83	94.11	50.57
CF40	57844072	55828494	8.7G	8.4G	0.03	97.72	93.85	51.18
CF49	130186356	19.53G	128268136	19.24G	0.02	98.13	94.74	49.25
CF50	125938148	18.89G	123352316	18.5G	0.03	97.91	94.23	48.78
CF51	129494924	19.42G	126746976	19.01G	0.02	98.29	95.2	50.26
CF52	133674248	20.05G	132047960	19.81G	0.03	97.8	93.92	49.28
CF53	124124958	18.62G	122131144	18.32G	0.02	98.18	94.95	49.84
CF54	136734206	20.51G	134567138	20.19G	0.02	98.28	95.05	48.7
CF55	121796346	18.27G	119485644	17.92G	0.02	98.03	94.37	48.06
CF56	105994278	15.9G	103968696	15.6G	0.02	98.18	94.87	49.96
CF57	144978038	21.75G	141459454	21.22G	0.02	98.14	94.85	49.05
CF58	103260746	15.49G	101791912	15.27G	0.02	98.15	94.77	49.22
CF59	105287902	15.79G	103477260	15.52G	0.02	98.33	95.14	48.74
CF60	103636386	15.55G	101347302	15.2G	0.02	98.17	94.92	49.44
CF61	111634568	16.75G	110130632	16.52G	0.02	98.16	94.95	49.59
CF62	121437840	18.22G	118638152	17.8G	0.02	98.28	95.2	48.26
CF63	126781774	19.02G	124212164	18.63G	0.02	98.15	94.86	48.78
CF64	117412116	17.61G	116086874	17.41G	0.03	97.79	93.92	49.35
CF65	114680340	17.2G	112753100	16.91G	0.03	97.87	93.92	47.66
CF66	107003740	16.05G	105040540	15.76G	0.02	98.25	95.07	49.45
CF67	132376352	19.86G	129333958	19.4G	0.02	97.99	94.34	49.57
CF68	112703376	16.91G	110318154	16.55G	0.02	98.07	94.53	49.75
CF69	139029978	20.85G	137243406	20.59G	0.02	97.98	94.29	48.89
CF70	129061590	19.36G	126929286	19.04G	0.02	97.97	94.42	49.73
CF71	148638522	22.3G	146130494	21.92G	0.02	97.99	94.42	50.39
CF72	113771242	17.07G	111701298	16.76G	0.02	98.11	94.5	47.56
CF73	102490078	15.37G	99320744	14.9G	0.02	98.36	95.4	50.18
CF74	134734580	20.21G	133063962	19.96G	0.02	98.12	94.56	48.62
CF75	123519276	18.53G	122022404	18.3G	0.02	98.16	94.83	48.66
CF76	108684786	16.3G	107047530	16.06G	0.02	98.19	94.93	47.64
CF77	111797670	16.77G	110338106	16.55G	0.02	98.25	95	49.86
CF78	117872532	17.68G	116248286	17.44G	0.02	98.1	94.75	49.22
CF79	157349588	23.6G	153019246	22.95G	0.02	98.36	95.41	49.61
CF80	131691064	19.75G	129677496	19.45G	0.02	98.31	95.16	50.16
CF81	108155580	16.22G	105731318	15.86G	0.02	98.17	94.78	49.91
CF82	119633394	17.95G	118151940	17.72G	0.02	98.22	94.96	49.34
CF83	99354390	14.9G	96154492	14.42G	0.02	98.2	94.96	49.66

CF84	118118574	17.72G	116394608	17.46G	0.02	98.3	95.13	49.3
CF85	131730510	19.76G	128954890	19.34G	0.02	98.21	94.97	49.92
CF86	110500334	16.58G	108534284	16.28G	0.02	98.27	95.17	49.6
CF87	124183640	18.63G	122852028	18.43G	0.02	98.24	94.95	49.43
CF88	109856690	16.48G	108546226	16.28G	0.02	98.02	94.56	49.51
CF89	128091422	19.21G	125912468	18.89G	0.02	98.22	94.97	49.32
CF90	102779268	15.42G	101356086	15.2G	0.02	98.28	95.06	50.47
CF91	114359956	17.15G	112784786	16.92G	0.02	98.18	94.91	49.46
CF92	125176470	18.78G	122892664	18.43G	0.02	98.04	94.39	48.81
CF93	114172562	17.13G	112514624	16.88G	0.02	98.23	95.08	50.24
CF94	126257964	18.94G	124601158	18.69G	0.02	98.38	95.31	49.35
CF95	116066710	17.41G	113950936	17.09G	0.03	97.85	93.75	48.33
CF96	121837010	18.28G	120275436	18.04G	0.02	98.04	94.58	49.22
CF97	109331750	16.4G	106747286	16.01G	0.02	98.19	94.97	49.8
CF98	106683862	16G	103487570	15.52G	0.02	98.3	95.3	50.2
CF99	130137366	19.52G	127043638	19.06G	0.02	98.41	95.46	48.46
CF100	126649592	19G	125035210	18.76G	0.02	98.36	95.32	49.95
CF101	127822104	19.17G	125756232	18.86G	0.02	98.22	95.07	50.28
CF102	122230852	18.33G	120592796	18.09G	0.02	98.15	94.83	48.88
CF103	100265244	15.04G	98492468	14.77G	0.02	98.08	94.71	49.91
CF104	107171418	16.08G	105642474	15.85G	0.02	98.21	94.93	49.31
CF105	105990724	15.9G	103017834	15.45G	0.02	98.2	94.95	49.86
CF106	125701770	18.86G	123848784	18.58G	0.02	98.2	94.94	49.52
CF107	154700346	23.21G	152695136	22.9G	0.02	98.14	94.87	48.37
CF108	121589670	18.24G	118955952	17.84G	0.02	98.27	95.19	49.61
CF109	119410232	17.91G	116738240	17.51G	0.02	98.33	95.26	49.71
CF110	105800452	15.87G	103647398	15.55G	0.02	98.13	94.83	48.85
CF111	136277238	20.44G	133300318	20.0G	0.02	98.13	94.92	49.17
CF112	110235858	16.54G	108487422	16.27G	0.02	98.27	95.09	49.88
CF113	115739382	17.36G	114007684	17.1G	0.02	98.24	95.02	50.35
CF114	119322040	17.9G	117548118	17.63G	0.03	97.83	94.03	50.16
CF115	124053644	18.61G	121274058	18.19G	0.02	98.19	94.92	49.42
CF116	105936512	15.89G	103360094	15.5G	0.02	98.32	95.39	50.84
CF117	141325264	21.2G	139035952	20.86G	0.02	98.1	94.67	49.37
CF118	158564728	23.78G	156608440	23.49G	0.02	98.22	94.94	49.49
CF119	129286870	19.39G	127517850	19.13G	0.02	98.15	94.8	49.42
CF120	121387206	18.21G	119788306	17.97G	0.02	98.19	94.86	50.13
CF121	127123616	19.07G	124726820	18.71G	0.02	98.29	95.25	49.88
CF122	123457980	18.52G	121658936	18.25G	0.02	98.24	94.93	48.63
CF123	126465416	18.97G	123595846	18.54G	0.02	98.3	95.27	49.65
CF124	120204580	18.03G	117681398	17.65G	0.02	98.34	95.23	49.21

CF125	104383880	15.66G	102708614	15.41G	0.02	98.16	94.6	47.86
CF126	104160004	15.62G	101442002	15.22G	0.02	98.38	95.35	50.01
CF127	124720222	18.71G	123295056	18.49G	0.02	98.28	95.07	49.47
CF128	129203150	19.38G	127341736	19.1G	0.02	98.04	94.52	49.07
CF129	115916188	17.39G	113259104	16.99G	0.02	98.28	95.16	50.1
CF130	117078586	17.56G	115134372	17.27G	0.02	98.04	94.52	49.73
CF131	125742596	18.86G	124031938	18.6G	0.02	98.14	94.77	50.1
CF132	115664172	17.35G	113892988	17.08G	0.02	98.24	95.01	50.01
CF133	115461970	17.32G	113276496	16.99G	0.02	98.23	94.95	48.8
CF134	108212814	16.23G	106693116	16.0G	0.02	98.34	95.26	49.82
CF135	142899736	21.43G	140495530	21.07G	0.02	98.09	94.73	49.73
CF136	118550508	17.78G	116771282	17.52G	0.02	98.31	95.13	49.94
CF137	103660946	15.55G	102173494	15.33G	0.02	98.35	95.16	48.57
CF138	103940890	15.59G	102537240	15.38G	0.03	97.79	93.89	49.75
CF139	120956176	18.14G	119294870	17.89G	0.02	98.08	94.44	48.66
CF140	140006968	21G	137979140	20.7G	0.02	98.11	94.66	50.01
CF141	124984878	18.75G	123390594	18.51G	0.02	98	94.42	49.71
CF142	115252666	17.29G	113550418	17.03G	0.02	98.29	95.09	49.95
CF143	102577706	15.39G	100342494	15.05G	0.02	98.27	95.09	49.41
CF144	109847532	16.48G	108353490	16.25G	0.02	98.15	94.82	49.61
CF145	108301852	16.25G	106295634	15.94G	0.02	98.32	95.24	50.3
CF146	119450768	17.92G	115714854	17.36G	0.02	98.15	94.78	50.17
CF147	119690586	17.95G	115824222	17.37G	0.02	98.38	95.49	50.14
CF148	105794486	15.87G	103601576	15.54G	0.02	98.01	94.4	49.99
CF149	106037338	15.91G	103998428	15.6G	0.02	98	94.4	50.56
CF150	106771176	16.02G	103598246	15.54G	0.02	98.2	94.97	51.14
CF151	138836504	20.83G	136243088	20.44G	0.02	98.09	94.61	50.17

Supplemental Table 4: Clinical information for bulk RNA-seq experiments. MM = mediastinal mass, PL = peripheral lymphadenopathy, SM = splenomegaly/splenic mass, HM = hepatomegaly/hepatic mass, U = unknown.

Sample name	Experiment group	MM	HC	PL	SM	HM
CF01	Cohort 1 CD4+ PTCL	U	Yes	Yes	U	U
CF02	Cohort 1 CD4+ PTCL	No	No	Yes	Yes	No
CF03	Cohort 1 CD4+ PTCL	No	Yes	Yes	No	No
CF04	Cohort 1 CD8+ PTCL	U	U	Yes	U	U
CF05	Cohort 1 CD4+ PTCL	U	U	Yes	Yes	U
CF06	Cohort 1 CD8+ PTCL	U	No	Yes	U	U
CF07	Cohort 1 CD4+ PTCL	U	U	Yes	U	U
CF08	Cohort 1 CD4-CD8- PTCL	U	U	Yes	U	U
CF09	Cohort 1 CD4-CD8- PTCL	U	No	Yes	U	U
CF10	Cohort 1 CD8+ PTCL	U	U	Yes	U	U

CF11	Cohort 1 CD8+ PTCL	U	Yes	Yes	U	U
CF12	Cohort 1 Control CD4+ nodal lymphocytes	No	No	No	No	No
CF13	Cohort 1 Control CD8+ nodal lymphocytes	No	No	No	No	No
CF14	Cohort 1 Control CD8+ nodal lymphocytes	No	No	No	No	No
CF15	Cohort 1 CD4+ PTCL	U	No	Yes	No	No
CF16	Cohort 1 Control CD8+ nodal lymphocytes	No	No	No	No	No
CF17	Cohort 1 Control CD4+ nodal lymphocytes	No	No	No	No	No
CF18	Cohort 1 CD4+ PTCL	Yes	Yes	Yes	No	No
CF19	Cohort 1 CD4+ PTCL	Yes	Yes	No	No	No
CF20	Cohort 1 CD4+ PTCL	U	U	Yes	Yes	U
CF21	Cohort 1 CD4+ PTCL	Yes	No	No	No	No
CF22	Cohort 1 CD4+ PTCL	U	U	Yes	U	U
CF23	Cohort 1 CD4+ PTCL	U	U	Yes	U	U
CF24	Cohort 1 CD4+ PTCL	Yes	Yes	U	U	U
CF25	Cohort 1 CD4+ PTCL	U	No	Yes	U	U
CF26	Cohort 1 CD4+ PTCL	Yes	Yes	Yes	No	U
CF27	Cohort 1 CD4+ PTCL	Yes	No	Yes	U	U
CF28	Cohort 1 CD4+ PTCL	U	No	Yes	U	U
CF29	Cohort 1 Control CD4+ nodal lymphocytes	No	No	No	No	No
CF30	Cohort 1 Control CD4+ nodal lymphocytes	No	No	No	No	No
CF31	Cohort 1 CD4+ PTCL	U	U	Yes	No	Yes
CF32	Cohort 1 CD4+ PTCL	No	No	Yes	U	U
CF33	Cohort 1 CD4+ PTCL	Yes	No	Yes	No	No
CF34	Cohort 1 CD4+ PTCL	U	No	Yes	U	U
CF35	Cohort 1 CD4+ PTCL	No	U	Yes	No	No
CF36	Cohort 1 CD4+ PTCL	U	No	Yes	No	No
CF37	Cohort 1 CD4+ PTCL	No	No	Yes	No	No
CF38	Cohort 1 CD4-CD8- PTCL	U	U	U	U	U
CF39	Cohort 1 CD4+ PTCL	U	U	U	U	U
CF40	Cohort 1 CD4+ PTCL	U	U	Yes	U	U
CF41	Thymus Control	No	No	No	No	No
CF42	Thymus Control	No	No	No	No	No
CF43	Thymus Control	No	No	No	No	No
CF44	Thymus Control	No	No	No	No	No
CF45	Thymus Control	No	No	No	No	No
CF46	Thymus Control	No	No	No	No	No
CF47	Thymus Control	No	No	No	No	No
CF48	Thymus Control	No	No	No	No	No
CF49	Cohort 2 CD4+ PTCL	U	U	Yes	U	U
CF50	Cohort 2 CD4+ PTCL	U	U	U	U	U
CF51	Cohort 2 CD4+ PTCL	No	No	Yes	Yes	Yes

CF52	Cohort 2 CD4+ PTCL	U	U	Yes	U	U
CF53	Cohort 2 CD4+ PTCL	No	No	Yes	Yes	Yes
CF54	Cohort 2 CD4+ PTCL	U	Yes	Yes	No	Yes
CF55	Cohort 2 CD4+ PTCL	U	U	Yes	Yes	U
CF56	Cohort 2 CD4+ PTCL	No	No	Yes	U	U
CF57	Cohort 2 CD4+ PTCL	U	U	U	U	U
CF58	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF59	Cohort 2 CD4+ PTCL	No	No	Yes	No	No
CF60	Cohort 2 CD4+ PTCL	U	No	Yes	U	U
CF61	Cohort 2 CD4+ PTCL	U	U	Yes	U	U
CF62	Cohort 2 CD4+ PTCL	U	U	Yes	U	U
CF63	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF64	Cohort 2 CD4+ PTCL	U	No	Yes	Yes	U
CF65	Cohort 2 CD4+ PTCL	Yes	Yes	Yes	U	U
CF66	Cohort 2 CD4+ PTCL	U	U	Yes	U	U
CF67	Cohort 2 CD4+ PTCL	U	U	Yes	U	U
CF68	Cohort 2 CD4+ PTCL	U	No	Yes	Yes	No
CF69	Cohort 2 CD4+ PTCL	No	No	Yes	U	U
CF70	Cohort 2 CD4+ PTCL	No	No	Yes	No	No
CF71	Cohort 2 CD4+ PTCL	Yes	No	Yes	No	No
CF72	Cohort 2 CD4+ PTCL	U	No	Yes	U	U
CF73	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF74	Cohort 2 CD4+ PTCL	U	No	Yes	No	No
CF75	Cohort 2 CD4+ PTCL	No	No	Yes	No	Yes
CF76	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF77	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF78	Cohort 2 CD4+ PTCL	No	Yes	Yes	Yes	No
CF79	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF80	Cohort 2 CD4+ PTCL	Yes	Yes	Yes	Yes	Yes
CF81	Cohort 2 CD4+ PTCL	Yes	Yes	Yes	Yes	Yes
CF82	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF83	Cohort 2 CD4+ PTCL	U	No	Yes	U	U
CF84	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF85	Cohort 2 CD4+ PTCL	Yes	No	Yes	Yes	No
CF86	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF87	Cohort 2 CD4+ PTCL	No	Yes	Yes	No	No
CF88	Cohort 2 CD4+ PTCL	U	No	Yes	U	U
CF89	Cohort 2 CD4+ PTCL	Yes	Yes	Yes	No	No
CF90	Cohort 2 CD4+ PTCL	No	No	Yes	U	U
CF91	Cohort 2 CD4+ PTCL	U	Yes	Yes	No	No
CF92	Cohort 2 CD4+ PTCL	No	No	Yes	Yes	Yes

CF93	Cohort 2 CD4+ PTCL	U	U	Yes	U	U
CF94	Cohort 2 CD4+ PTCL	No	Yes	Yes	Yes	No
CF95	Cohort 2 CD4+ PTCL	Yes	Yes	No	No	Yes
CF96	Cohort 2 CD4+ PTCL	No	No	Yes	No	No
CF97	Cohort 2 CD4+ PTCL	U	U	Yes	U	U
CF98	Cohort 2 CD4+ PTCL	U	No	Yes	No	No
CF99	Cohort 2 CD4+ PTCL	Yes	Yes	Yes	U	U
CF100	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF101	Cohort 2 CD4+ PTCL	U	No	Yes	U	U
CF102	Cohort 2 CD4+ PTCL	U	U	Yes	No	No
CF103	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF104	Cohort 2 CD4+ PTCL	U	No	Yes	U	U
CF105	Cohort 2 CD4+ PTCL	No	No	Yes	No	No
CF106	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF107	Cohort 2 CD4+ PTCL	No	No	Yes	No	No
CF108	Cohort 2 CD4+ PTCL	Yes	No	Yes	No	No
CF109	Cohort 2 CD4+ PTCL	U	U	Yes	U	U
CF110	Cohort 2 CD4+ PTCL	No	No	Yes	U	U
CF111	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF112	Cohort 2 CD4+ PTCL	Yes	Yes	Yes	No	Yes
CF113	Cohort 2 CD4+ PTCL	No	No	Yes	U	U
CF114	Cohort 2 CD4+ PTCL	U	No	Yes	U	U
CF115	Cohort 2 CD4+ PTCL	No	No	Yes	U	U
CF116	Cohort 2 CD4+ PTCL	U	No	Yes	U	U
CF117	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF118	Cohort 2 CD4+ PTCL	Yes	No	Yes	Yes	No
CF119	Cohort 2 CD4+ PTCL	U	U	U	U	U
CF120	Cohort 2 CD4+ PTCL	U	No	Yes	U	U
CF121	Cohort 2 CD4+ PTCL	U	U	Yes	U	U
CF122	Cohort 2 CD4+ PTCL	No	Yes	Yes	No	No
CF123	Cohort 2 CD4+ PTCL	No	Yes	Yes	U	U
CF124	Cohort 2 CD4+ PTCL	No	Yes	Yes	No	Yes
CF125	Cohort 2 CD4+ PTCL	U	U	Yes	U	U
CF126	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF127	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF128	Cohort 2 CD4+ PTCL	U	U	Yes	No	No
CF129	Cohort 2 CD4+ PTCL	U	No	Yes	U	U
CF130	Cohort 2 CD4+ PTCL	No	Yes	Yes	Yes	Yes
CF131	Cohort 2 CD4+ PTCL	U	No	Yes	U	U
CF132	Cohort 2 CD4+ PTCL	Yes	Yes	Yes	U	U
CF133	Cohort 2 CD4+ PTCL	No	No	Yes	No	No

CF134	Cohort 2 CD4+ PTCL	No	Yes	Yes	Yes	No
CF135	Cohort 2 CD4+ PTCL	Yes	Yes	Yes	U	U
CF136	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF137	Cohort 2 CD4+ PTCL	Yes	No	Yes	U	U
CF138	Cohort 2 CD4+ PTCL	U	No	Yes	U	U
CF139	Cohort 2 CD4+ PTCL	No	Yes	Yes	Yes	No
CF140	Cohort 2 CD4+ PTCL	U	No	Yes	No	No
CF141	Cohort 2 CD4+ PTCL	U	No	Yes	U	U
CF142	Cohort 2 CD4+ PTCL	No	Yes	Yes	Yes	U
CF143	Cohort 2 CD4+ PTCL	U	Yes	Yes	No	No
CF144	Cohort 2 CD4+ PTCL	U	Yes	Yes	U	U
CF145	Cohort 2 Control CD4+ nodal lymphocytes	No	No	No	No	No
CF146	Cohort 2 Control CD4+ nodal lymphocytes	No	No	No	No	No
CF147	Cohort 2 Control CD4+ nodal lymphocytes	No	No	No	No	No
CF148	Cohort 2 Control CD4+ thymocytes	No	No	No	No	No
CF149	Cohort 2 Control CD4+ nodal lymphocytes	No	No	No	No	No
CF150	Cohort 2 Control CD4+ nodal lymphocytes	No	No	No	No	No
CF151	Cohort 2 Control CD4+ thymocytes	No	No	No	No	No

Supplemental Table 5: Flow cytometric features of control CD4+ nodal lymphocytes. Flow cytometry data is not available for samples CF149 and CF150.

	CF12 (Dog #1)	**CF12 (Dog #2)**	**CF29**	**CF30**	**CF145**	**CF146**	**CF147**
% CD4+	23.43	33.65	35.31	33.74	35.22	23.92	32.8
% CD8+	7.45	15.09	12.11	8.77	10.55	13.26	7.57
% DN	7.42	7.6	5.84	3.98	5.81	11.96	9.99
%CD25+ (CD4 T-cells)	7.2	11.27	6.01	4.7	8.57	4.08	4.18
%CD25+ (CD8 T-cells)	1.4	0.64	0.75	0.5	0.8	13.26	0.84
%CD25+ (DN cells)	8.36	14.77	9.45	7.82	13.84	11.96	9.99
MHC CL II MFI (CD4 T-cells)	4.37	10.44	9.44	13.01	12.58	16.43	12.61
CD5 MFI (CD4 T-cells)	40.69	42.22	38.29	65.17	N/A	N/A	N/A
CD3 MFI (CD4 T-cells)	6.43	5.84	6.43	9.25	8.82	8.47	8.39
CD8 MFI (CD4 T-cells)	0.15	0.16	0.15	0.12	N/A	N/A	N/A
CD4 MFI (CD4 T-cells)	5.84	6.44	6.41	7.41	7	8.46	7.42
CD5 MFI (CD8 T-cells)	23.69	23.64	23.3	40.97	N/A	N/A	N/A
CD3 MFI (CD8 T-cells)	5.14	4.66	5.19	7.35	7.11	6.12	6.53
CD8 MFI (CD8 T-cells)	13.51	14.97	13.17	15.67	14.52	16.01	15.93
CD4 MFI (CD8 T-cells)	0.17	0.17	0.17	0.18	N/A	N/A	N/A

Supplemental Table 6: Flow cytometric features of control CD4+ thymocytes.

	CF41	**CF42**	**CF43**	**CF44**	**CF45**	**CF46**	**CF47**	**CF48**
% CD4 SP	12.98	14.58	5.43	4.3	7.42	6.87	9.89	17.64
% CD8 SP	14.18	12.06	9.39	9.05	12.42	8.05	8.2	14.25
% DN	14.49	14.34	11.87	15.53	10.43	15.52	16.34	19.23
% DP	58.35	59.02	72.77	71.13	69.72	69.56	65.57	48.88
%CD25+ (CD4 SP cells)	4.65	2.69	3.42	3.36	1.11	1.45	N/A	N/A
%CD25+ (CD8 SP cells)	0.08	0.06	0.04	0.03	0.02	0	N/A	N/A
%CD25+ (DN cells)	2	1.57	0.59	0.71	0.82	0.33	N/A	N/A
CD25 %CD25+ (DP cells)	0.09	0.11	0.05	0.06	0	0.05	N/A	N/A
MHC CL II MFI (all CD4+ cells)	1.2	0.74	0.14	0.15	0.21	0.19	3.43	5.5
MHC CL II MFI (CD4 SP cells)	N/A	N/A	N/A	N/A	N/A	N/A	10.77	6.06
MHC CL II MFI (CD8 SP cells)	N/A	N/A	N/A	N/A	N/A	N/A	9.79	10.05
MHC CL II MFI (DN cells)	N/A	N/A	N/A	N/A	N/A	N/A	17.59	8.77
MHC CL II MFI (DP cells)	N/A	N/A	N/A	N/A	N/A	N/A	2.98	4.38
CD5 MFI (CD4 SP cells)	121.33	125.66	65.81	52.53	104.41	64.34	25.64	17.56
CD3 MFI (CD4 SP cells)	12.66	13.35	10.85	8.93	12.58	9.77	10.03	11.82
CD8 MFI (CD4 SP cells)	0.86	0.88	0.56	0.49	0.34	0.6	0.19	0.08
CD4 MFI (CD4 SP cells)	3.09	3.31	2.67	2.54	2.74	2.83	7.13	3.97

Supplemental Table 7: MultiQC RNA-seq QA/QC report for the CanFam3.1 alignment of Cohort 1 and Cohort 2.

Sample Name	featureCounts: % Assigned reads	featureCounts: Assigned reads (millions)	STAR: % Uniquely mapped reads	Star: Uniquely mapped reads (millions)
CF01	63.9%	16.3	94.2%	23.1
CF02	60.3%	15.2	93.5%	22.5
CF03	56.2%	15.5	94.1%	24.7
CF04	56.3%	17.7	94.5%	28.9
CF05	53.9%	17.9	94.3%	30.2
CF06	62.9%	17.6	94.0%	25.4
CF07	59.9%	16.5	94.8%	25.3
CF08	65.5%	19.3	94.0%	26.7
CF09	59.8%	19.9	93.7%	29.6
CF10	63.2%	26.9	93.2%	37.8
CF11	60.9%	20.8	91.7%	30
CF12	62.1%	22.6	94.4%	33.5
CF13	60.0%	17.1	94.4%	26.1
CF14	59.4%	17.4	94.3%	26.5
CF15	66.0%	15.9	93.4%	21.6
CF16	58.6%	15.5	94.1%	24
CF17	60.5%	15.5	93.8%	23.2

CF18	57.5%	15.2	93.1%	23
CF19	64.8%	19	94.5%	26.6
CF20	61.7%	20.6	93.7%	30.2
CF21	63.9%	16.6	93.5%	23.2
CF22	61.3%	20.2	93.2%	29
CF23	64.8%	19	93.5%	25.8
CF24	58.8%	17.1	93.0%	25.5
CF25	58.6%	18.2	93.8%	27.6
CF26	55.1%	17.5	93.4%	28.1
CF27	63.6%	22	94.2%	31.4
CF28	63.6%	18.6	94.7%	26.9
CF29	64.6%	22.7	93.6%	31.5
CF30	62.5%	22.7	94.3%	33.1
CF31	63.5%	20.1	93.3%	28.1
CF32	63.0%	22	92.9%	30.9
CF33	63.6%	15.5	94.1%	22.1
CF34	62.2%	18.6	94.5%	27.3
CF35	61.4%	21.6	92.2%	30.2
CF36	61.9%	20.2	94.5%	30.3
CF37	55.6%	18.1	93.6%	29.1
CF38	60.7%	24.7	92.4%	35.1
CF39	58.4%	15.5	93.6%	23.5
CF40	62.4%	18.7	93.7%	26.8
CF49	61.1%	80.4	95.4%	61.6
CF50	57.1%	72.3	95.6%	59.7
CF51	61.8%	81.4	95.1%	61.1
CF52	59.4%	80.2	95.4%	63.2
CF53	62.2%	77.8	95.6%	58.9
CF54	61.6%	85.1	96.0%	65.1
CF55	66.8%	82.1	95.9%	58.0
CF56	62.9%	67.5	95.5%	50.2
CF57	55.9%	82.4	95.1%	68.3
CF58	58.4%	60.7	95.8%	49.1
CF59	65.0%	69.0	95.9%	50.1
CF60	57.8%	60.7	94.9%	48.8
CF61	59.7%	67.3	95.5%	52.8
CF62	62.7%	76.9	95.7%	57.6
CF63	62.6%	79.9	95.8%	60.2
CF64	60.5%	71.3	94.9%	55.2
CF65	59.9%	69.3	95.8%	54.5
CF66	60.5%	65.5	95.5%	50.7

CF67	61.7%	82.7	95.1%	62.5
CF68	64.2%	73.3	95.1%	53.2
CF69	65.2%	91.3	95.5%	65.8
CF70	58.4%	75.4	95.8%	61.3
CF71	63.2%	94.9	95.1%	70.1
CF72	68.1%	78.1	96.1%	54.2
CF73	63.4%	66.2	95.1%	48.3
CF74	69.5%	94.2	95.4%	63.7
CF75	60.1%	74.8	95.4%	58.5
CF76	68.8%	75.8	95.5%	51.4
CF77	64.9%	73.3	95.4%	52.9
CF78	60.0%	71.1	95.7%	55.9
CF79	61.7%	99.1	95.1%	74.2
CF80	65.5%	87.1	95.2%	62.2
CF81	67.9%	74.9	95.5%	51.3
CF82	68.3%	82.5	95.8%	56.9
CF83	59.8%	60.1	95.5%	47.1
CF84	63.8%	76.1	95.7%	56.1
CF85	64.1%	85.6	95.2%	62.2
CF86	60.5%	67.3	95.9%	52.5
CF87	68.8%	86.4	95.6%	59.0
CF88	65.4%	72.9	95.3%	51.9
CF89	64.7%	84.9	94.8%	60.3
CF90	66.1%	69.3	94.6%	48.3
CF91	64.7%	75.6	94.9%	53.8
CF92	67.7%	85.0	96.0%	59.7
CF93	64.5%	75.5	95.0%	53.8
CF94	67.5%	86.6	95.3%	59.8
CF95	66.0%	77.2	95.6%	55.1
CF96	57.7%	71.5	95.1%	57.5
CF97	60.0%	66.4	95.2%	51.6
CF98	61.5%	66.7	94.9%	50.2
CF99	66.0%	86.7	95.7%	61.8
CF100	64.5%	82.0	95.3%	59.9
CF101	62.2%	81.6	94.1%	59.6
CF102	57.0%	70.0	95.5%	57.9
CF103	61.8%	63.2	95.1%	47.2
CF104	60.3%	65.4	95.3%	50.7
CF105	61.3%	66.6	94.5%	49.7
CF106	61.6%	79.0	95.3%	59.4
CF107	63.7%	99.1	95.4%	73.1

CF108	63.4%	78.0	95.3%	57.5
CF109	63.2%	77.7	95.0%	56.3
CF110	54.3%	58.1	95.5%	50.0
CF111	55.6%	77.3	94.9%	64.1
CF112	64.6%	72.0	95.4%	52.2
CF113	65.8%	77.1	95.0%	54.6
CF114	60.6%	73.8	94.2%	55.7
CF115	64.3%	81.4	95.4%	58.7
CF116	60.1%	64.5	95.2%	50.0
CF117	62.9%	90.7	94.8%	66.4
CF118	63.6%	102.6	94.9%	74.6
CF119	61.8%	81.1	95.3%	61.1
CF120	65.2%	80.0	95.2%	57.3
CF121	63.2%	81.1	95.1%	60.0
CF122	62.5%	77.9	96.1%	58.9
CF123	58.6%	76.1	94.5%	59.3
CF124	68.2%	83.0	95.6%	57.0
CF125	67.8%	71.3	95.9%	49.7
CF126	67.7%	71.9	94.4%	48.8
CF127	65.0%	82.3	95.5%	59.1
CF128	61.7%	81.6	95.1%	60.9
CF129	64.0%	75.7	95.0%	54.6
CF130	66.0%	80.3	94.1%	54.7
CF131	70.6%	89.4	95.4%	59.4
CF132	66.0%	77.8	95.3%	54.7
CF133	61.1%	71.4	95.7%	54.8
CF134	65.2%	71.2	95.2%	51.1
CF135	58.9%	86.5	94.2%	66.7
CF136	66.2%	79.6	95.4%	56.2
CF137	66.0%	69.7	95.7%	49.3
CF138	62.8%	65.9	95.3%	49.1
CF139	69.2%	83.8	95.1%	56.9
CF140	64.0%	90.2	95.8%	66.6
CF141	63.7%	79.9	95.7%	59.3
CF142	65.6%	76.6	95.4%	54.6
CF143	65.7%	69.4	94.6%	48.2
CF144	63.0%	70.0	95.4%	52.0
CF145	61.7%	67.9	95.2%	51.2
CF146	61.5%	75.2	94.9%	56.2
CF147	61.5%	75.5	94.9%	56.4
CF148	60.6%	66.0	94.7%	49.7

CF149	62.7%	68.1	94.7%	49.8
CF150	63.7%	69.4	94.7%	50.2
CF151	61.5%	87.5	94.4%	65.0

Supplemental Table 8: MultiQC RNA-seq QA/QC report for the ROS Cfam 1.0 alignment of Cohort 2.

Sample Name	featureCounts: % Assigned reads	featureCounts: Assigned reads (millions)	STAR: % Uniquely mapped reads	Star: Uniquely mapped reads (millions)
CF49	70.2%	89.7	93.7%	60.5
CF50	64.1%	78.7	93.2%	58.2
CF51	71.7%	93.2	94.9%	61.0
CF52	69.1%	91.7	94.2%	62.4
CF53	71.5%	87.4	94.2%	58.0
CF54	72.3%	98.6	95.4%	64.7
CF55	63.8%	69.4	85.3%	51.6
CF56	69.9%	72.9	93.7%	49.3
CF57	63.8%	91.1	93.9%	67.4
CF58	69.1%	71.1	95.5%	48.9
CF59	71.1%	72.6	92.6%	48.4
CF60	65.7%	67.6	94.3%	48.4
CF61	67.3%	73.8	93.9%	52.0
CF62	63.5%	72.7	89.9%	54.1
CF63	70.4%	87.2	93.4%	58.7
CF64	69.2%	80.1	93.9%	54.6
CF65	69.1%	78.5	94.6%	53.9
CF66	69.0%	73.5	94.7%	50.3
CF67	68.1%	88.0	92.4%	60.7
CF68	73.0%	81.7	93.9%	52.5
CF69	67.6%	89.8	90.6%	62.4
CF70	68.2%	87.1	95.3%	61.0
CF71	74.3%	109.9	94.6%	69.7
CF72	61.2%	62.9	86.0%	48.4
CF73	70.9%	71.5	93.1%	47.4
CF74	69.5%	86.8	87.7%	58.6
CF75	69.8%	85.3	94.2%	57.7
CF76	60.0%	57.5	83.5%	44.9
CF77	71.8%	78.7	93.5%	51.9
CF78	70.0%	80.5	93.3%	54.5
CF79	70.9%	111.0	94.3%	73.6
CF80	74.7%	97.7	94.0%	61.4
CF81	75.6%	81.2	93.7%	50.3
CF82	77.3%	91.2	93.9%	55.7

CF83	67.2%	65.8	94.0%	46.3
CF84	72.1%	84.0	93.9%	55.1
CF85	73.9%	96.8	94.4%	61.7
CF86	69.5%	75.8	94.7%	51.9
CF87	76.1%	92.9	93.1%	57.4
CF88	71.1%	74.9	90.7%	49.4
CF89	73.3%	92.0	92.9%	59.1
CF90	74.2%	76.2	93.4%	47.7
CF91	71.5%	80.4	93.0%	52.7
CF92	71.5%	84.8	90.3%	56.2
CF93	74.3%	84.5	93.9%	53.2
CF94	72.0%	89.5	92.5%	58.0
CF95	76.5%	88.0	94.0%	54.2
CF96	66.0%	79.9	94.4%	57.0
CF97	68.0%	73.0	93.7%	50.8
CF98	70.4%	74.8	94.4%	50.0
CF99	67.6%	83.3	90.4%	58.4
CF100	72.4%	89.4	93.1%	58.4
CF101	70.4%	88.5	93.1%	59.0
CF102	66.1%	79.6	94.3%	57.1
CF103	68.4%	67.3	93.5%	46.5
CF104	68.7%	72.8	93.9%	49.9
CF105	68.4%	71.5	92.6%	48.7
CF106	69.3%	85.2	93.2%	58.1
CF107	67.1%	99.8	91.3%	70.0
CF108	71.6%	85.6	93.6%	56.4
CF109	70.9%	84.3	93.9%	55.6
CF110	64.2%	67.4	94.9%	49.8
CF111	64.0%	86.8	94.5%	63.8
CF112	75.2%	83.2	95.2%	52.1
CF113	73.2%	84.0	93.8%	53.9
CF114	69.2%	81.4	92.8%	54.9
CF115	73.5%	90.7	94.4%	58.1
CF116	69.9%	73.4	94.6%	49.7
CF117	70.4%	97.8	92.5%	64.8
CF118	70.2%	109.3	93.0%	73.2
CF119	71.2%	91.7	94.7%	60.7
CF120	72.0%	85.5	92.6%	55.8
CF121	67.1%	82.2	91.7%	57.8
CF122	71.3%	87.0	94.4%	57.9
CF123	67.8%	84.5	93.5%	58.6

CF124	74.4%	86.4	91.8%	54.8
CF125	74.1%	74.8	92.1%	47.8
CF126	75.0%	76.3	91.9%	47.5
CF127	74.1%	92.0	94.1%	58.3
CF128	67.6%	84.5	91.5%	58.6
CF129	73.9%	84.7	93.9%	54.0
CF130	70.0%	80.2	90.3%	52.4
CF131	72.6%	83.6	87.2%	54.3
CF132	76.0%	87.8	95.0%	54.5
CF133	68.2%	77.6	94.1%	53.9
CF134	75.6%	80.5	93.7%	50.3
CF135	65.4%	91.0	92.5%	65.5
CF136	71.9%	82.1	91.6%	53.9
CF137	72.5%	73.3	92.0%	47.4
CF138	74.1%	77.1	95.0%	48.9
CF139	65.4%	73.1	88.1%	52.7
CF140	73.5%	102.0	94.4%	65.6
CF141	73.6%	90.5	93.9%	58.2
CF142	75.1%	85.7	94.0%	53.8
CF143	74.6%	74.7	92.1%	46.9
CF144	74.5%	81.9	95.4%	52.0
CF145	71.4%	78.1	95.3%	51.2
CF146	71.6%	86.2	95.0%	56.3
CF147	72.3%	87.2	95.1%	56.5
CF148	72.3%	77.5	95.6%	50.2
CF149	74.2%	79.5	95.4%	50.2
CF150	74.8%	80.9	95.5%	50.6
CF151	72.8%	101.9	95.1%	65.5

Supplemental Table 9: MultiQC RNA-seq QA/QC report for the CanFam4 alignment of Cohort 2.

Sample Name	featureCounts: % Assigned reads	featureCounts: Assigned reads (millions)	STAR: % Uniquely mapped reads	Star: Uniquely mapped reads (millions)
CF49	69.7%	93.4	95.8%	61.8
CF50	64.0%	82.6	96.0%	59.9
CF51	71.6%	95.4	96.0%	61.7
CF52	69.0%	95.1	96.3%	63.8
CF53	71.5%	90.8	96.4%	59.3
CF54	72.3%	101.1	96.8%	65.7
CF55	65.6%	82.5	95.0%	57.4
CF56	69.9%	76.4	95.9%	50.4
CF57	63.9%	94.4	95.4%	68.5

CF58	68.7%	72.6	96.5%	49.4
CF59	72.5%	78.3	96.3%	50.4
CF60	65.4%	69.4	95.9%	49.3
CF61	67.3%	77.0	95.9%	53.1
CF62	65.7%	82.2	95.3%	57.4
CF63	71.1%	92.4	96.2%	60.5
CF64	70.3%	84.1	95.5%	55.5
CF65	69.9%	81.8	96.4%	54.9
CF66	69.0%	75.8	96.1%	51.0
CF67	67.7%	92.4	95.8%	62.9
CF68	73.3%	85.3	95.9%	53.6
CF69	69.4%	99.7	95.1%	65.5
CF70	68.1%	89.5	96.5%	61.7
CF71	74.5%	113.5	95.9%	70.7
CF72	69.4%	81.3	95.5%	53.8
CF73	71.3%	75.5	95.3%	48.4
CF74	72.8%	100.7	95.2%	63.6
CF75	70.0%	88.7	96.0%	58.8
CF76	66.3%	74.6	93.5%	50.3
CF77	71.9%	82.6	95.8%	53.1
CF78	69.8%	84.1	96.4%	56.3
CF79	70.2%	113.5	95.7%	74.7
CF80	75.2%	101.9	96.1%	62.9
CF81	76.4%	85.0	96.1%	51.6
CF82	77.7%	95.4	96.4%	57.2
CF83	68.3%	69.5	96.0%	47.3
CF84	73.2%	88.7	96.5%	56.6
CF85	74.5%	100.3	95.9%	62.7
CF86	69.5%	78.5	96.6%	53.0
CF87	77.4%	98.9	96.1%	59.3
CF88	70.5%	79.6	95.2%	51.8
CF89	74.0%	96.9	95.3%	60.6
CF90	73.2%	78.3	95.4%	48.7
CF91	72.0%	84.8	95.4%	54.1
CF92	73.3%	94.4	96.1%	59.7
CF93	73.1%	86.7	95.2%	53.9
CF94	74.5%	97.3	95.8%	60.1
CF95	78.0%	92.4	96.3%	55.6
CF96	66.2%	82.3	96.0%	58.0
CF97	67.9%	75.8	96.0%	52.0
CF98	70.2%	76.5	95.8%	50.7

CF99	72.2%	96.0	95.5%	61.7
CF100	72.9%	94.5	95.9%	60.2
CF101	71.0%	92.6	94.7%	60.0
CF102	66.1%	82.3	96.4%	58.4
CF103	71.1%	72.5	95.2%	47.3
CF104	68.7%	75.4	96.2%	51.2
CF105	70.1%	76.1	94.4%	49.6
CF106	70.5%	89.9	96.0%	59.9
CF107	69.5%	110.4	95.7%	73.4
CF108	72.3%	90.0	95.7%	57.7
CF109	70.9%	87.5	95.4%	56.6
CF110	64.4%	69.1	96.0%	50.3
CF111	63.7%	89.3	95.5%	64.5
CF112	74.9%	84.9	96.3%	52.7
CF113	72.7%	87.2	95.8%	55.0
CF114	68.6%	83.8	94.9%	56.1
CF115	74.3%	94.7	95.9%	59.0
CF116	69.5%	75.5	96.1%	50.5
CF117	70.4%	102.6	95.1%	66.7
CF118	70.4%	115.1	95.5%	75.1
CF119	71.0%	94.1	95.7%	61.3
CF120	73.4%	92.1	95.5%	57.5
CF121	69.4%	90.3	95.1%	60.0
CF122	72.1%	91.2	96.6%	59.2
CF123	67.1%	87.4	94.7%	59.4
CF124	75.5%	93.0	96.0%	57.3
CF125	75.8%	81.3	95.8%	49.7
CF126	76.2%	81.2	95.0%	49.1
CF127	74.5%	95.6	96.5%	59.7
CF128	69.6%	92.3	94.9%	60.8
CF129	73.4%	87.0	95.6%	55.0
CF130	71.5%	87.7	94.2%	54.7
CF131	72.4%	94.1	94.5%	58.8
CF132	76.8%	90.9	96.1%	55.2
CF133	68.6%	81.3	95.9%	54.9
CF134	76.4%	84.5	96.1%	51.6
CF135	65.7%	95.3	94.5%	66.9
CF136	71.9%	87.9	95.7%	56.3
CF137	73.1%	78.9	96.0%	49.4
CF138	73.6%	78.8	96.4%	49.7
CF139	71.8%	88.6	95.2%	57.0

CF140	73.4%	105.7	96.4%	67.0
CF141	73.3%	94.0	96.2%	59.6
CF142	75.5%	89.5	96.0%	54.9
CF143	75.0%	78.3	94.7%	48.2
CF144	74.3%	83.2	96.3%	52.5
CF145	71.7%	79.7	96.5%	51.8
CF146	71.2%	87.6	95.8%	56.8
CF147	71.6%	88.3	95.8%	56.9
CF148	72.5%	78.4	95.7%	50.3
CF149	74.1%	80.3	95.8%	50.4
CF150	74.3%	81.9	96.1%	50.9
CF151	72.6%	103.4	95.4%	65.7

Supplemental Table 10: MultiQC RNA-seq QA/QC report for the CanFam6 alignment of Cohort 2.

Sample Name	featureCounts: % Assigned reads	featureCounts: Assigned reads (millions)	STAR: % Uniquely mapped reads	Star: Uniquely mapped reads (millions)
CF49	72.7%	89.2	90.5%	58.4
CF50	66.5%	77.9	89.7%	56.0
CF51	73.8%	93.1	93.3%	60.0
CF52	71.1%	91.0	91.6%	60.7
CF53	74.4%	86.9	91.3%	56.2
CF54	74.3%	98.1	93.4%	63.4
CF55	71.7%	67.9	74.5%	45.1
CF56	72.9%	72.5	90.2%	47.4
CF57	66.2%	90.8	91.0%	65.4
CF58	70.9%	70.7	93.6%	47.9
CF59	75.4%	71.7	86.9%	45.5
CF60	67.6%	67.4	92.2%	47.4
CF61	70.2%	73.4	90.8%	50.2
CF62	69.1%	71.3	81.6%	49.1
CF63	73.8%	86.4	89.2%	56.1
CF64	72.0%	79.1	89.9%	52.3
CF65	71.5%	77.8	91.4%	52.0
CF66	71.5%	73.3	92.4%	49.1
CF67	70.8%	87.2	88.9%	58.4
CF68	75.8%	81.1	90.6%	50.7
CF69	73.4%	88.2	82.8%	57.0
CF70	70.2%	86.8	93.3%	59.7
CF71	77.0%	109.5	92.1%	67.9
CF72	72.6%	59.8	69.0%	38.9
CF73	74.7%	71.2	89.0%	45.3

CF74	77.4%	84.6	77.1%	51.5
CF75	72.3%	84.7	91.3%	55.9
CF76	72.4%	55.6	67.0%	36.1
CF77	75.0%	78.0	89.5%	49.6
CF78	72.7%	79.9	89.9%	52.6
CF79	73.1%	110.6	92.1%	71.9
CF80	77.6%	96.9	90.7%	59.3
CF81	79.1%	80.6	90.0%	48.3
CF82	80.7%	90.7	90.5%	53.7
CF83	70.4%	65.3	90.0%	44.4
CF84	75.4%	83.4	90.0%	52.8
CF85	76.8%	96.8	91.8%	60.0
CF86	71.8%	75.4	92.2%	50.5
CF87	80.4%	92.2	88.2%	54.4
CF88	75.3%	74.6	86.6%	47.1
CF89	77.2%	91.8	88.9%	56.5
CF90	76.3%	75.7	91.1%	46.5
CF91	75.1%	80.2	89.1%	50.5
CF92	77.9%	83.4	82.2%	51.1
CF93	77.1%	84.7	92.0%	52.1
CF94	76.9%	88.0	85.9%	53.9
CF95	80.0%	87.5	90.6%	52.3
CF96	67.9%	79.5	92.3%	55.7
CF97	70.5%	72.6	90.8%	49.3
CF98	72.5%	74.6	92.6%	49.0
CF99	75.0%	81.4	80.2%	51.8
CF100	76.1%	88.5	88.5%	55.6
CF101	74.1%	88.4	89.2%	56.5
CF102	67.9%	79.1	92.0%	55.7
CF103	73.1%	66.6	87.4%	43.5
CF104	70.8%	72.3	91.6%	48.7
CF105	72.8%	71.6	87.9%	46.2
CF106	72.7%	84.7	89.3%	55.6
CF107	72.1%	98.4	84.7%	64.9
CF108	75.1%	85.3	89.9%	54.2
CF109	74.0%	84.1	90.8%	53.8
CF110	66.3%	67.4	93.3%	48.9
CF111	66.4%	87.0	92.5%	62.5
CF112	76.9%	82.7	93.3%	51.1
CF113	76.0%	83.7	91.1%	52.3
CF114	71.0%	80.9	90.5%	53.5

CF115	76.8%	90.4	91.1%	56.1
CF116	72.2%	73.2	92.5%	48.6
CF117	73.7%	97.4	88.9%	62.4
CF118	73.7%	108.7	89.0%	70.0
CF119	73.7%	91.3	92.3%	59.2
CF120	76.5%	84.4	86.8%	52.3
CF121	72.5%	81.4	85.2%	53.7
CF122	74.0%	86.2	90.9%	55.7
CF123	70.9%	84.7	90.8%	56.9
CF124	79.2%	85.7	86.4%	51.6
CF125	79.3%	73.6	85.3%	44.2
CF126	79.8%	75.9	86.7%	44.9
CF127	76.7%	91.3	91.2%	56.5
CF128	72.8%	84.2	85.5%	54.8
CF129	76.8%	84.7	91.6%	52.7
CF130	75.5%	79.8	83.7%	48.6
CF131	79.7%	82.4	78.8%	49.1
CF132	79.0%	87.5	92.0%	52.8
CF133	71.3%	77.1	90.6%	51.9
CF134	78.8%	80.0	89.9%	48.3
CF135	68.9%	91.1	88.9%	62.9
CF136	76.0%	81.5	86.7%	51.1
CF137	77.1%	72.8	87.0%	44.8
CF138	75.9%	76.6	93.2%	48.0
CF139	74.8%	69.7	73.5%	44.0
CF140	76.2%	101.6	91.9%	63.9
CF141	76.6%	89.9	90.8%	56.2
CF142	78.5%	85.3	90.6%	51.9
CF143	78.7%	74.7	88.1%	44.9
CF144	76.3%	81.5	93.8%	51.1
CF145	73.4%	78.0	94.2%	50.6
CF146	73.6%	86.3	94.0%	55.7
CF147	74.2%	87.4	94.2%	56.0
CF148	74.4%	77.8	94.7%	49.7
CF149	76.0%	79.7	94.7%	49.8
CF150	76.4%	80.9	94.6%	50.1
CF151	74.9%	102.3	94.0%	64.7

Appendix 2: Supplementary Material for Chapter Two

Data availability

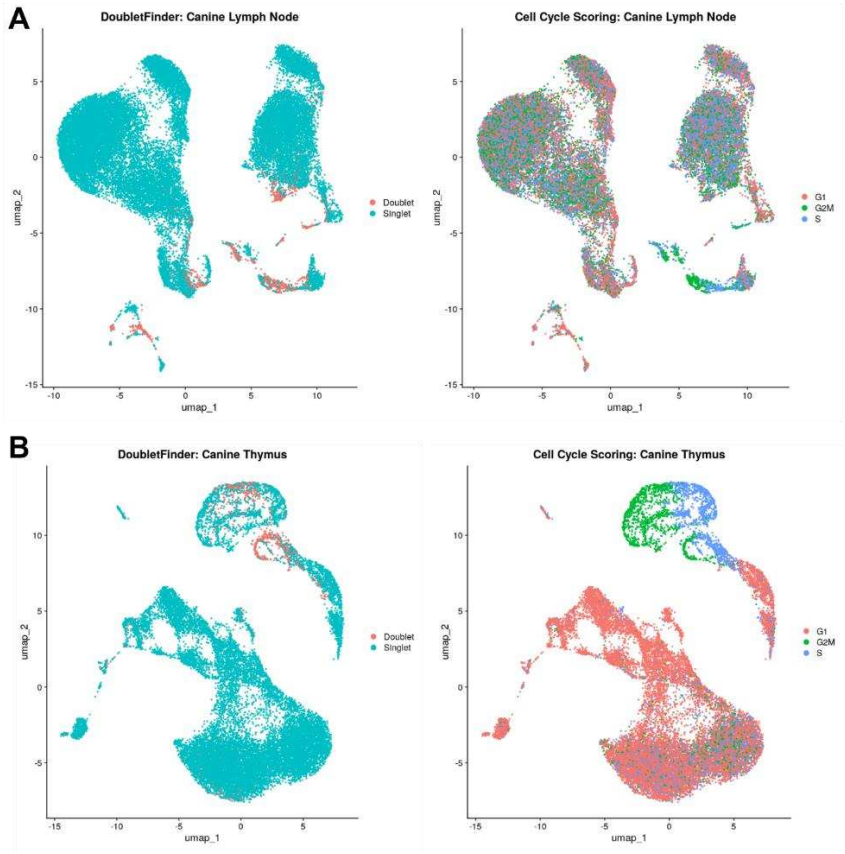
Normalized bulk RNA-seq read counts, DESeq2 results for the top 500 differentially expressed genes between canine $CD4^+$ PTCL compared to control $CD4^+$ lymphocytes, and full GSEA and GSVA results for Cohort 1 are available as supplementary data through our publication in *BMC Cancer*: https://doi.org/10.1186/s12885-023-11762-w.

Scripts and software packages used for processing, aligning, and tabulating bulk RNA-seq reads and performing differential expression analysis, hierarchical clustering, GSEA, and GSVA are documented in a Github repository: https://github.com/pirateyoho/CaninePTCLRNASeq. Scripts and software packages used for single-cell RNA-seq data analysis and pseudotemporal trajectory analysis of the normal canine thymus and lymph node are documented in an additional Github repository: https://github.com/pirateyoho/CanineSingleCellRNASeq.

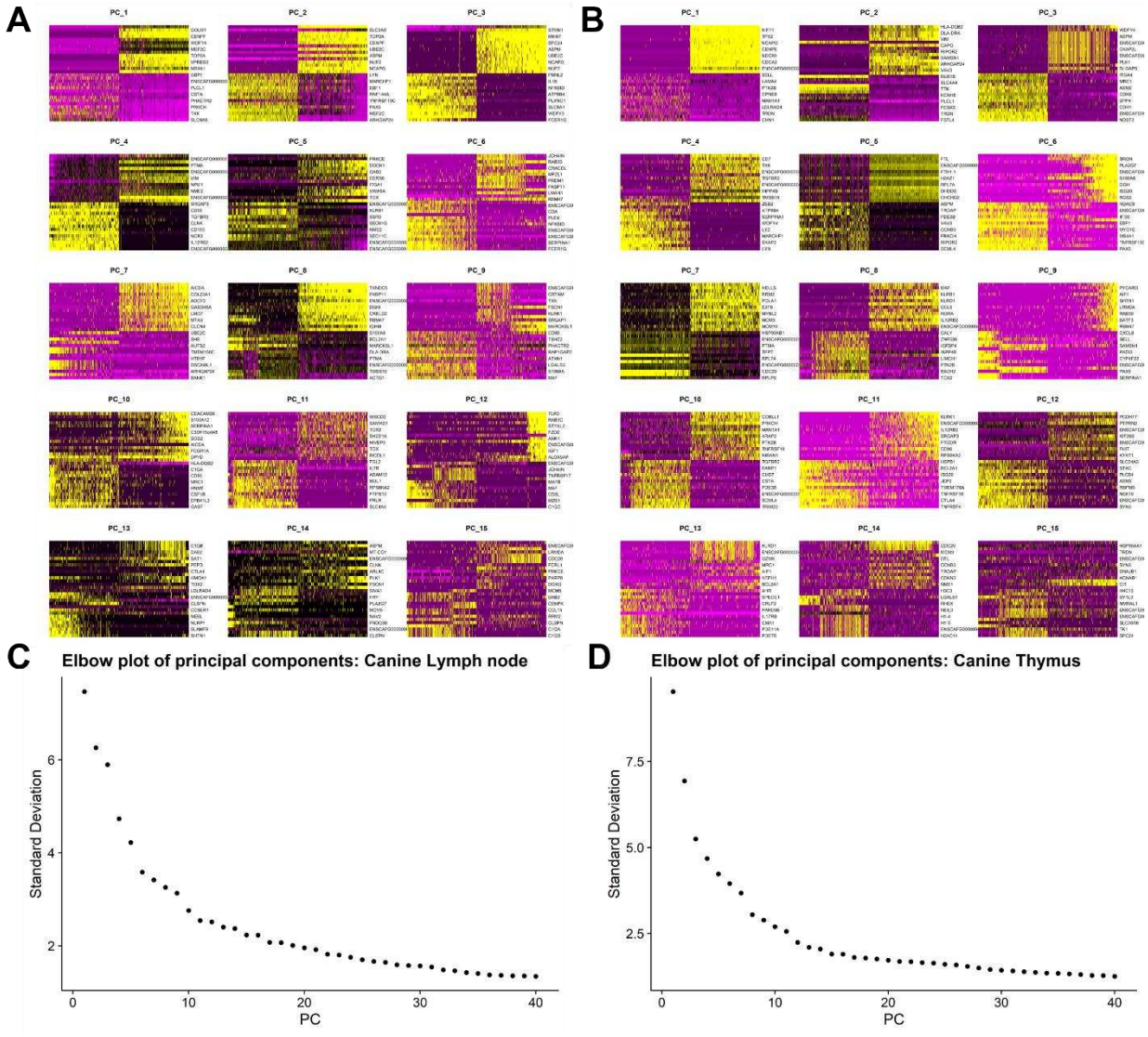
Source code with results for Chapter 2 figures is also available on RPubs:

Figure(s)	Source code
Figure 19A, Figure 20B, Figure 21A	https://rpubs.com/eileenowens/deseq2_cohort1
Figure 19B, Figure 20C, Figure 21B	https://rpubs.com/eileenowens/deseq2_cohort2
Figure 20A, Figure 22	https://rpubs.com/eileenowens/CD4PTCL_GSEA_GSVA
Figure 23A	https://rpubs.com/eileenowens/singleCellNormAndClust_LN
Figure 23B,D,F	https://rpubs.com/eileenowens/LNAnnot3
Figure 23C	https://rpubs.com/eileenowens/LNAnnot1
Figure 23E	https://rpubs.com/eileenowens/LNAnnot2
Figure 24A	https://rpubs.com/eileenowens/singleCellNormalizationAndClusteringThymus
Figure 24B,D,F	https://rpubs.com/eileenowens/ThymAnnot3
Figure 24C	https://rpubs.com/eileenowens/thymusAnnot1
Figure 24E	https://rpubs.com/eileenowens/ThymAnnot2

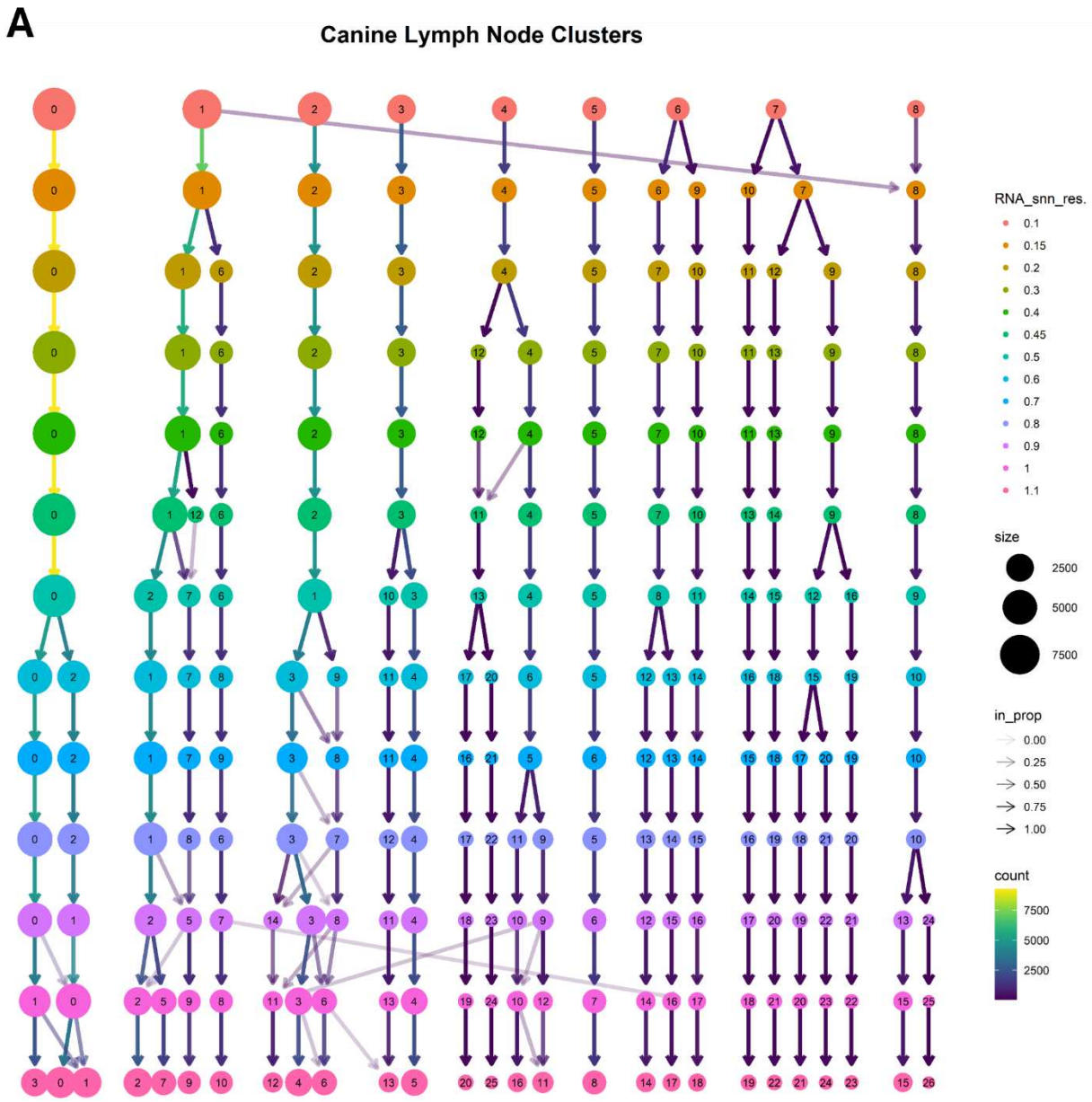
Figure 25A-B	https://rpubs.com/eileenowens/single-cell-t-cell-integration
Figure 25C	https://rpubs.com/eileenowens/merged-sc-marker-exp
Figure 26 through Figure 36	https://rpubs.com/eileenowens/merged-sc-top-marker-exp https://rpubs.com/eileenowens/merged-scdata-gsea-T https://rpubs.com/eileenowens/merged-sc-gsea-thym
Figure 37, Figure 38A	https://rpubs.com/eileenowens/merged-scdata-gsea-T
Figure 38B, Figure 39	https://rpubs.com/eileenowens/pseudotime-monocle3
Figure 40, Figure 41	https://rpubs.com/eileenowens/pseudotime-gene-exp
Figure 42, Figure 43	https://rpubs.com/eileenowens/ptcl-to-seurat
Figure 44, Figure 45	https://rpubs.com/eileenowens/ptcl-to-monocle3
Supplemental Figure 5	https://rpubs.com/eileenowens/doubletfinder
Supplemental Figure 6A,C **Supplemental Figure 7A**	https://rpubs.com/eileenowens/singleCellNormAndClust_LN
Supplemental Figure 6B,D **Supplemental Figure 7B**	https://rpubs.com/eileenowens/singleCellNormalizationAndClusteringThymus
Supplemental Figure 8	https://rpubs.com/eileenowens/seurat_QC



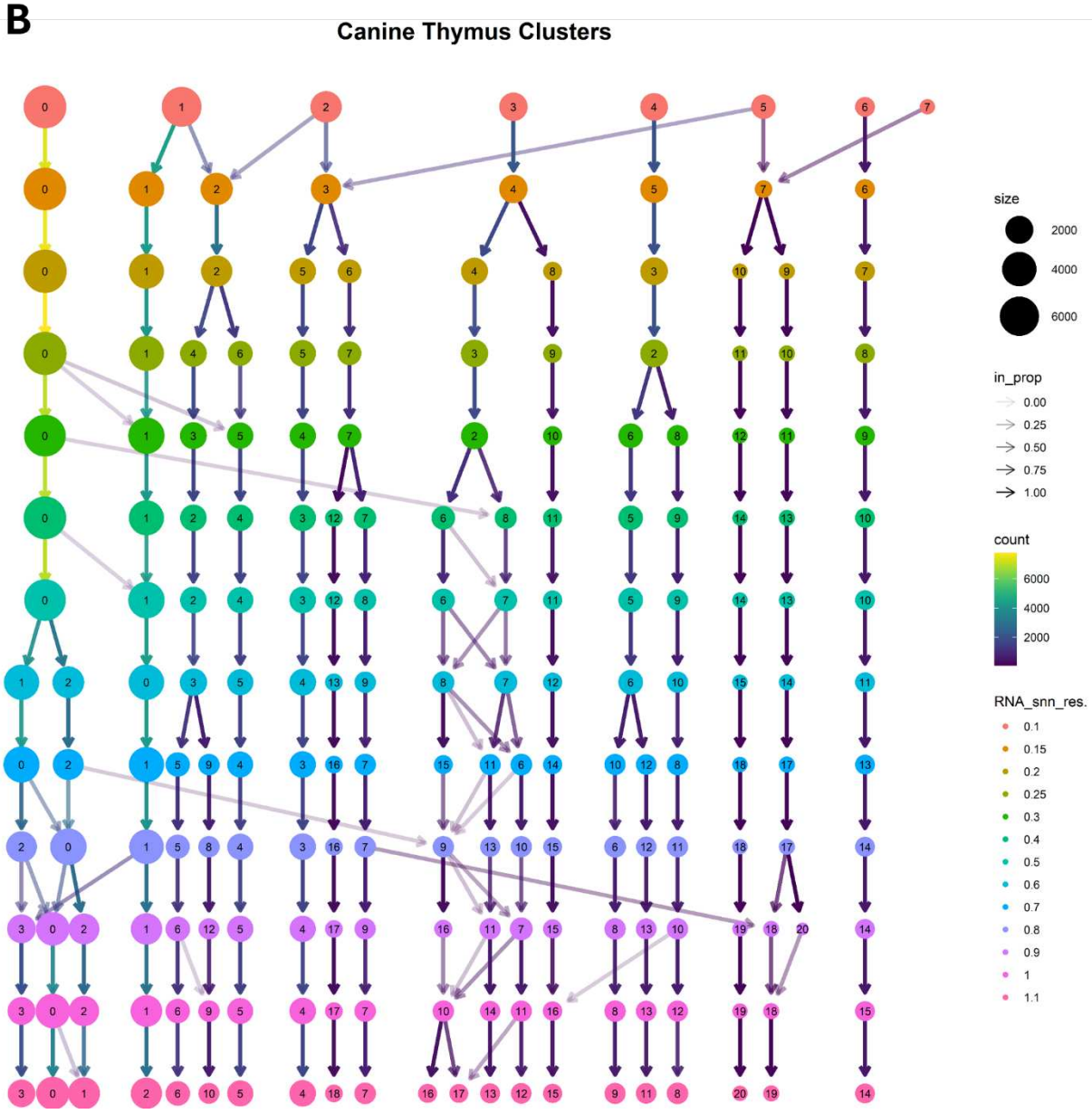
Supplemental Figure 5: DoubletFinder results. The vast majority of doublets predicted by DoubletFinder in single-cell RNA-seq data from normal canine lymph node (**A**) and thymus (**B**) coincided with cells enriched for G2M and S phase genes of the cell cycle. The high amount of transcript in these cells was therefore deemed biologically relevant rather than artifact, and doublet calls were not filtered from our lymph node or thymus single-cell data sets.

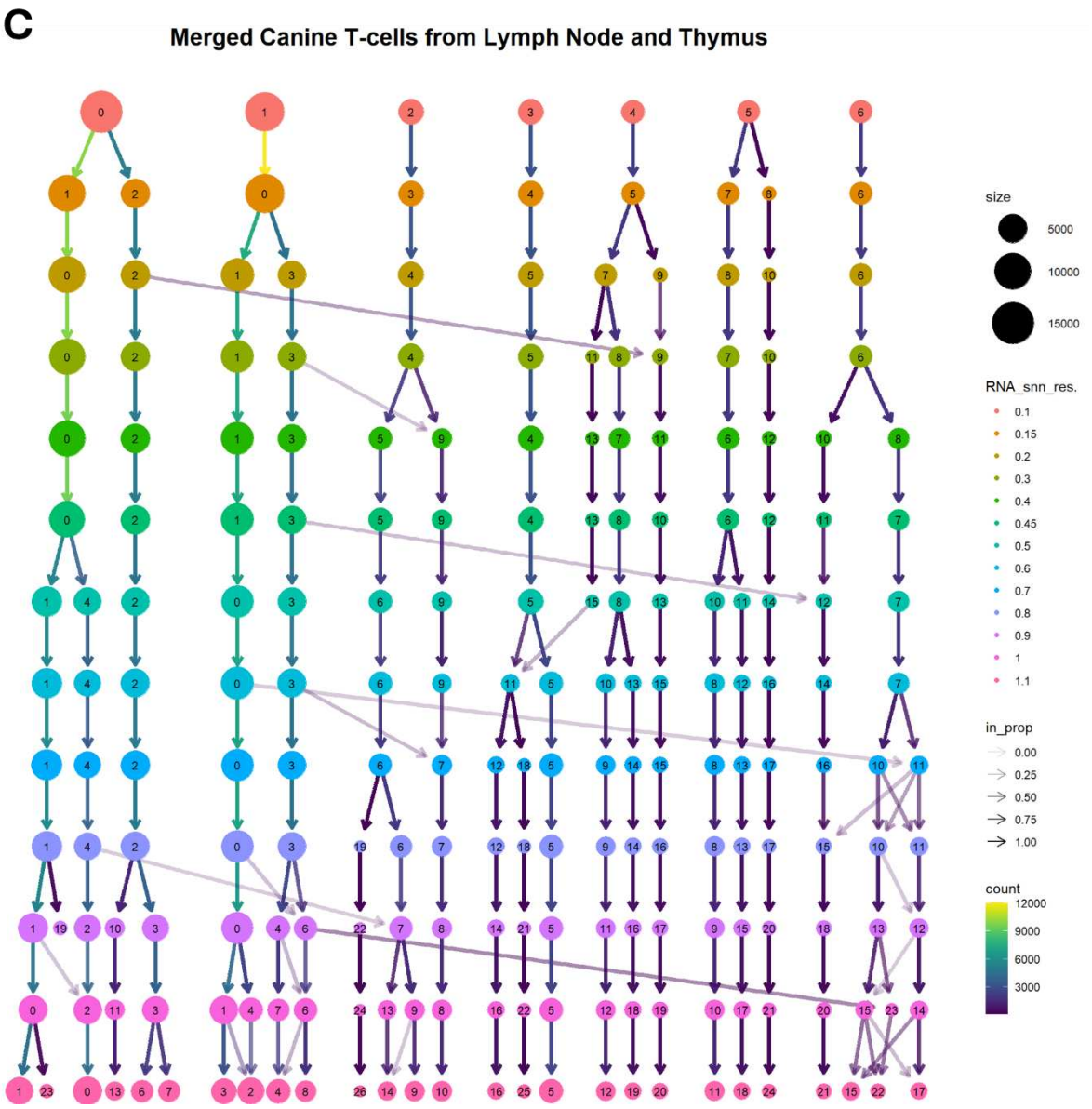


Supplemental Figure 6: Principal component analysis of single-cell RNA-seq data. The first 15 principal components for canine lymph node (**A**) and thymus (**B**) single-cell RNA-seq data. The number of principal components to use in linear dimensionality reduction was determined by ranking of principal components based on percentage of variance explained by the "elbow" in an elbow plot. For both the lymph node (**C**) and thymus (**D**), the majority of the true signal appeared to be captured in the first 10 principal components.

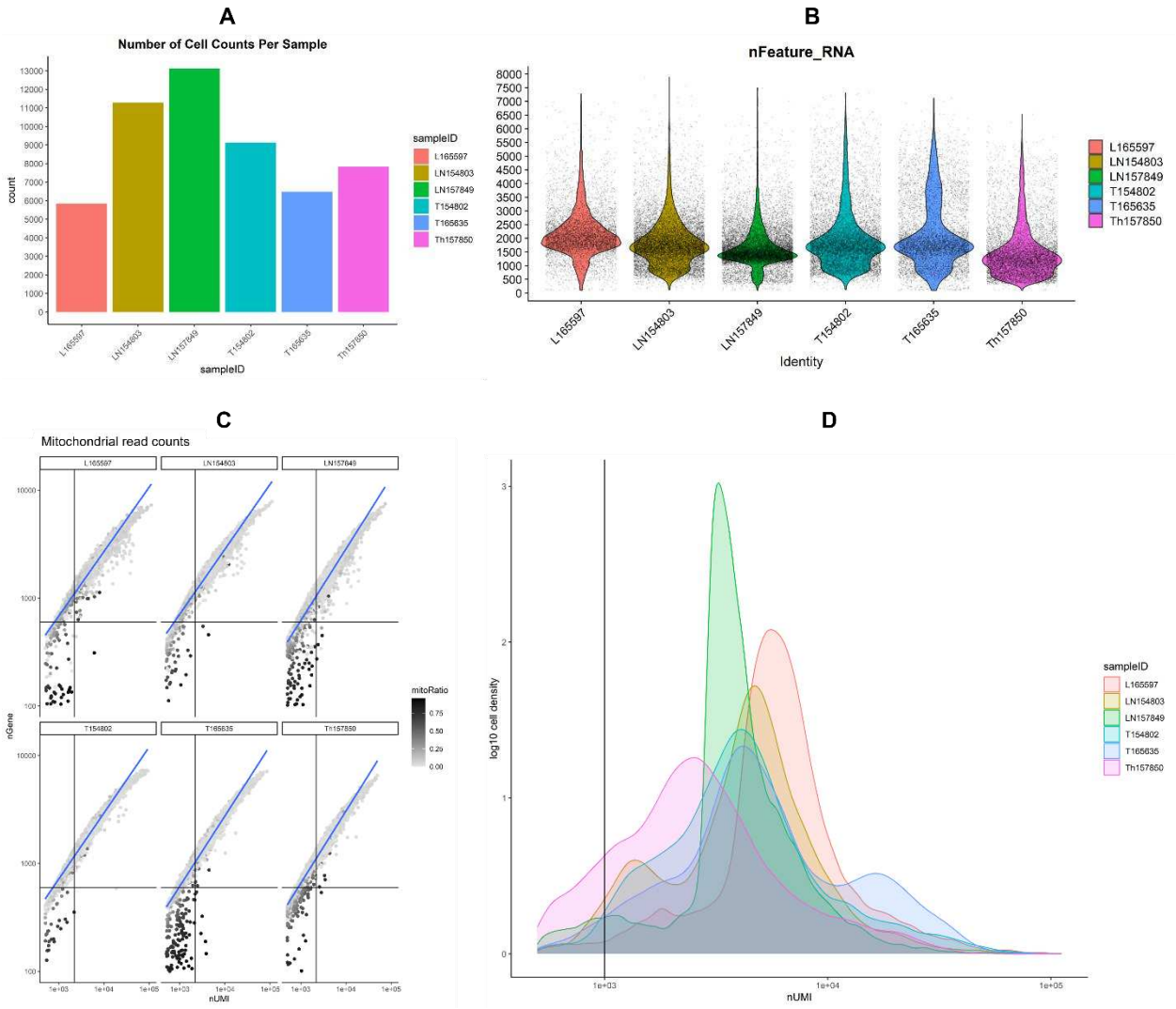

A
Canine Lymph Node Clusters
RNA_snn_res.
0.1
0.15
0.2
0.3
0.4
0.45
0.5
0.6
0.7
0.8
0.9
1
1.1
size
2500
5000
7500
in_prop
0.00
0.25
0.50
0.75
1.00
count
7500
5000
2500

B

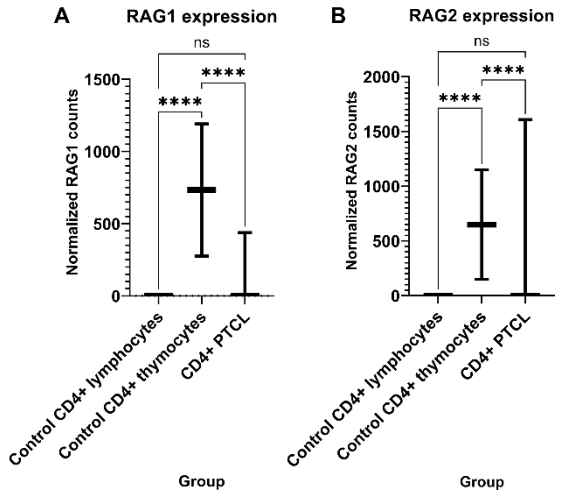

Canine Thymus Clusters
size
2000
4000
6000
in_prop
0.00
0.25
0.50
0.75
1.00
count
6000
4000
2000
RNA_snn_res.
0.1
0.15
0.2
0.25
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
1.1



Supplemental Figure 7: Analysis of cluster resolutions with clustree. Determining the optimal resolution and number of clusters for single-cell data that strikes an appropriate balance between over- and under-clustering the data can be challenging. Clustering trees are one method for comparing the stability of clusterings at a range of resolutions. The node size is related to the number of cells that make up each cluster, and node color corresponds to clustering resolution. Arrows are colored according to the number of samples they represent, and arrow transparency corresponds to incoming node proportion. "Messier" branches in the tree, where nodes have multiple incoming edges, indicates over-clustering of the data. Resolutions 0.15-0.3 and 0.5 for the lymph node (**A**), 0.15, 0.2, and 0.3 for the thymus, (**B**) and 0.1, 0.15, and 0.4 for the merged lymph node and thymus (**C**) appeared the most stable.



Supplemental Figure 8: Quality metrics for single-cell RNA-seq data of normal canine thymus and lymph node. (**A**) Single-cell RNA-seq data was collected from a range of 6,480-9,130 cells per sample of canine thymus and 5,844-13,126 cells per sample of canine lymph node. (**B**) The average number of genes (nFeature_RNA) detected per cell prior to filtering was 1,839 for thymus samples and 1,835 for lymph node samples. (**C**) A plot of the number of genes per cell versus the number of UMIs per cell, colored by the fraction of mitochondrial reads. Mitochondrial read fractions are high in particularly low count cells with few detected genes in both lymph node (top) and thymus (bottom) samples, consistent with dying cells. (**D**) A histogram demonstrating the distribution of UMIs (transcripts) detected per cell. UMIs per cell should generally be at least over 500, with most cells having 1000 UMIs or more (denoted by the vertical line).



Supplemental Figure 9: Expression of *RAG1* (**A**) and *RAG2* (**B**) genes is significantly downregulated in CD4⁺ PTCL compared to control CD4⁺ thymocytes.

Supplemental Table 11: The original reference atlas cluster corresponding to each gene set used for annotation of canine thymus and lymph node single-cell RNA-seq clusters via module scoring and GSEA.

Assigned gene set name	Reference atlas	Original cluster name
Park_Double_Negative	https://cells.ucsc.edu/?ds=fetal-thymus+all	DN
Park_Early_Thymic_Progenitor	https://cells.ucsc.edu/?ds=fetal-thymus+all	ETP
Park_Double_Positive	https://cells.ucsc.edu/?ds=fetal-thymus+all	DP
Park_CD4_Single_Positive	https://cells.ucsc.edu/?ds=fetal-thymus+all	CD4+T
Park_CD8_Single_Positive	https://cells.ucsc.edu/?ds=fetal-thymus+all	CD8+T
Park_T_Agonist	https://cells.ucsc.edu/?ds=fetal-thymus+all	T(agonist)
Park_Innate_Lymphoid_Cell_Type_3	https://cells.ucsc.edu/?ds=fetal-thymus+all	ILC3
Park_CD4_T_Memory	https://cells.ucsc.edu/?ds=fetal-thymus+all	CD4+Tmem
Park_Treg	https://cells.ucsc.edu/?ds=fetal-thymus+all	Treg
Park_NK	https://cells.ucsc.edu/?ds=fetal-thymus+all	NK
Park_CD8_T_memory	https://cells.ucsc.edu/?ds=fetal-thymus+all	CD8+Tmem
Park_NKT	https://cells.ucsc.edu/?ds=fetal-thymus+all	NKT
Park_Megakaryocyte	https://cells.ucsc.edu/?ds=fetal-thymus+hsc	Mgk
Park_Erythrocyte	https://cells.ucsc.edu/?ds=fetal-thymus+hsc	Ery
Park_Double_Negative_T	https://cells.ucsc.edu/?ds=fetal-thymus+hsc	T_DN
Park_Mast_Cell	https://cells.ucsc.edu/?ds=fetal-thymus+hsc	Mast
Park_MEMP	https://cells.ucsc.edu/?ds=fetal-thymus+hsc	MEMP
Park_HSC	https://cells.ucsc.edu/?ds=fetal-thymus+hsc	HSC
Park_NMP	https://cells.ucsc.edu/?ds=fetal-thymus+hsc	NMP
Park_proB	https://cells.ucsc.edu/?ds=fetal-thymus+hsc	B_pro/pre
Park_Plasmacytoid_DC	https://cells.ucsc.edu/?ds=fetal-thymus+hsc	pDC

Park_Early_Thymic_Progenitor_II	https://cells.ucsc.edu/?ds=fetal-thymus+hsc	ETP
Park_DN_Early	https://cells.ucsc.edu/?ds=fetal-thymus+tcells	DN(early)
Park_DN_Proliferating	https://cells.ucsc.edu/?ds=fetal-thymus+tcells	DN(P)
Park_DP_Proliferating	https://cells.ucsc.edu/?ds=fetal-thymus+tcells	DP(P)
Park_DP_Quiescent	https://cells.ucsc.edu/?ds=fetal-thymus+tcells	DP(Q)
Park_T_Agonist_II	https://cells.ucsc.edu/?ds=fetal-thymus+tcells	T(agonist)
Park_Treg_Differentiating	https://cells.ucsc.edu/?ds=fetal-thymus+tcells	Treg(diff)
Park_Treg_II	https://cells.ucsc.edu/?ds=fetal-thymus+tcells	Treg
Park_CD4_Single_Positive_II	https://cells.ucsc.edu/?ds=fetal-thymus+tcells	CD4+T
Park_CD8_Single_Positive_II	https://cells.ucsc.edu/?ds=fetal-thymus+tcells	CD8+T
Park_Th17	https://cells.ucsc.edu/?ds=fetal-thymus+tcells	Th17
Park_NKT_II	https://cells.ucsc.edu/?ds=fetal-thymus+tcells	NKT
Park_DN_Quiescent	https://cells.ucsc.edu/?ds=fetal-thymus+tcells	DN(Q)
Ammons_Immature_B_cell	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	Immature B cell
Ammons_Naïve_B_cell	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	Naïve B cell
Ammons_CD8_gd_T_cell	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	CD8 gd T cell
Ammons_Cycling_T_cell	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	Cycling T cell
Ammons_CD34_Unclassified	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	CD34+ Unclassified
Ammons_Unclassified_DC	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	Unclassified DC
Ammons_Pre_DC	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	Pre-DC
Ammons_PMN_MDSC	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	PMN-MDSC
Ammons_Plasmacytoid_DC	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	Plasmacytoid DC
Ammons_Plasma_cell	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	Plasma cell
Ammons_NKT_cell	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	NK T cell
Ammons_DN_T_cell	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	DN T cell
Ammons_CD8_Naive	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	CD8+ Naïve

Ammons_CD4_Naive	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	CD4+ Naïve
Ammons_CD4_TCM	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	CD4+ TCM
Ammons_CD4_Treg	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	CD4+ T reg
Ammons_CD4_IFN_signature	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	CD4+, IFN signature
Ammons_CD4_TEM	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	CD4+ TEM
Ammons_CD4_TEM_Th1_like	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	CD4+ TEM, Th1 like
Ammons_CD4_TEM_Th17_like	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	CD4+ TEM, Th17 like
Ammons_CD4_TEM_Th2_like	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	CD4+ TEM, Th2 like
Ammons_CD8_Memory	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	CD8+ Memory
Ammons_CD8_Effector	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	CD8+ Effector
Ammons_NK_cell	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	NK cell
Ammons_gd_T_cell	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	gd T cell
Ammons_Myeloid_c_DC2	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	Myeloid cDC2
Ammons_Monocyte_CD4_neg	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	Monocyte, CD4-
Ammons_Monocyte_IFN_signature	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	Monocyte, IFN signature
Ammons_Monocyte_CD4_pos	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	Monocyte, CD4+
Ammons_M_MDSC	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	M-MDSC
Ammons_Basophil	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	Basophil
Ammons_Neutrophil	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	Neutrophil
Ammons_Eosinophil	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	Eosinophil
Ammons_Class_switched_B_cell	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	Class switched B cell
Ammons_Activated_B_cell	https://cells.ucsc.edu/?ds=canine-leukocyte-atlas+healthy	Activated B cell
Conde_Trm_gut_CD8	https://cells.ucsc.edu/?ds=pan-immune+global	Trm_gut_CD8
Conde_TnaiveCM_CD4	https://cells.ucsc.edu/?ds=pan-immune+global	Tnaive/CM_CD4
Conde_Tregs	https://cells.ucsc.edu/?ds=pan-immune+global	Tregs
Conde_Pro_B	https://cells.ucsc.edu/?ds=pan-immune+global	Pro-B

Conde_Naïve_B	https://cells.ucsc.edu/?ds=pan-immune+global	Naïve B cells
Conde_Pre_B	https://cells.ucsc.edu/?ds=pan-immune+global	Pre-B
Conde_Plasma_Cell	https://cells.ucsc.edu/?ds=pan-immune+global	Plasma cells
Conde_Plasmablasts	https://cells.ucsc.edu/?ds=pan-immune+global	Plasmablasts
Conde_Memory_B	https://cells.ucsc.edu/?ds=pan-immune+global	Memory B cells
Conde_Erythroid	https://cells.ucsc.edu/?ds=pan-immune+global	Erythroid
Conde_Plasmacytoid_DC	https://cells.ucsc.edu/?ds=pan-immune+global	pDC
Conde_Megakaryocytes	https://cells.ucsc.edu/?ds=pan-immune+global	Megakaryocytes
Conde_Mast_cells	https://cells.ucsc.edu/?ds=pan-immune+global	Mast cells
Conde_ILC3	https://cells.ucsc.edu/?ds=pan-immune+global	ILC3
Conde_Cycling_T_NK	https://cells.ucsc.edu/?ds=pan-immune+global	Cycling T&NK
Conde_DC_1	https://cells.ucsc.edu/?ds=pan-immune+global	DC1
Conde_DC_2	https://cells.ucsc.edu/?ds=pan-immune+global	DC2
Conde_Classical_Monocyte	https://cells.ucsc.edu/?ds=pan-immune+global	Classical monocytes
Conde_Cycling	https://cells.ucsc.edu/?ds=pan-immune+global	Cycling
Conde_Progenitor	https://cells.ucsc.edu/?ds=pan-immune+global	Progenitor
Conde_CD8_TEMRA	https://cells.ucsc.edu/?ds=pan-immune+global	Tem/emra_CD8
Conde_Migratory_DC	https://cells.ucsc.edu/?ds=pan-immune+global	migDC
Conde_Trm_Tgd	https://cells.ucsc.edu/?ds=pan-immune+global	Trm_Tgd
Conde_NK_CD56_Bright_CD16_Negative	https://cells.ucsc.edu/?ds=pan-immune+global	NK_CD56bright_CD16-
Conde_NK_CD16_Positive	https://cells.ucsc.edu/?ds=pan-immune+global	NK_CD16+
Conde_Tgd_CRTAM_Positive	https://cells.ucsc.edu/?ds=pan-immune+global	Tgd_CRTAM+
Conde_Tmem_CD8	https://cells.ucsc.edu/?ds=pan-immune+global	Trm/em_CD8
Conde_MAIT	https://cells.ucsc.edu/?ds=pan-immune+global	MAIT
Conde_T_naive_CM_CD4_activated	https://cells.ucsc.edu/?ds=pan-immune+global	Tnaive/CM_CD4_activated
Conde_Trm_Th1_Th17	https://cells.ucsc.edu/?ds=pan-immune+global	Trm_Th1/Th17

Conde_T_CD4_CD8	https://cells.ucsc.edu/?ds=pan-immune+global	T_CD4/CD8
Conde_T_naive_CM_CD8	https://cells.ucsc.edu/?ds=pan-immune+global	Tnaive/CM_CD8
Conde_TEM_CD4	https://cells.ucsc.edu/?ds=pan-immune+global	Teffector/EM_CD4
Conde_Tfh	https://cells.ucsc.edu/?ds=pan-immune+global	Tfh

Supplemental Table 12: Gene sets from the Molecular Signatures Database (MSigDB) used for annotation of single-cell clusters by gene set enrichment analysis.

MSigDB Category	**Gene set name**
C5 Ontology	GOBP REGULATION OF THYMOCYTE APOPTOTIC PROCESS
C5 Ontology	GOBP T HELPER 17 CELL DIFFERENTIATION
C5 Ontology	GOBP T HELPER 2 CELL DIFFERENTIATION
C5 Ontology	GOBP THYMOCYTE MIGRATION
C7 Immunologic Signature	GSE11924 TFH VS TH1 CD4 TCELL UP
C7 Immunologic Signature	GSE11924 TFH VS TH17 CD4 TCELL UP
C7 Immunologic Signature	GSE11924 TFH VS TH2 CD4 TCELL UP
C7 Immunologic Signature	GSE11924 TH1 VS TH17 CD4 TCELL UP
C7 Immunologic Signature	GSE11924 TH1 VS TH2 CD4 TCELL UP
C7 Immunologic Signature	GSE11924 TH2 VS TH17 CD4 TCELL UP
C7 Immunologic Signature	GSE13493 CD4INTCD8POS VS CD8POS THYMOCYTE UP
C7 Immunologic Signature	GSE13493 DP VS CD4INTCD8POS THYMOCYTE UP
C7 Immunologic Signature	GSE13493 DP VS CD8POS THYMOCYTE UP
C7 Immunologic Signature	GSE14308 INDUCED VS NATURAL TREG UP
C7 Immunologic Signature	GSE14308 NAIVE CD4 TCELL VS INDUCED TREG UP
C7 Immunologic Signature	GSE14308 NAIVE CD4 TCELL VS NATURAL TREG UP
C7 Immunologic Signature	GSE14308 TH1 VS INDUCED TREG UP
C7 Immunologic Signature	GSE14308 TH1 VS NAIVE CD4 TCELL UP
C7 Immunologic Signature	GSE14308 TH1 VS NATURAL TREG UP
C7 Immunologic Signature	GSE14308 TH1 VS TH17 UP

C7 Immunologic Signature	GSE14308 TH17 VS INDUCED TREG UP
C7 Immunologic Signature	GSE14308 TH17 VS NAIVE CD4 TCELL UP
C7 Immunologic Signature	GSE14308 TH17 VS NATURAL TREG UP
C7 Immunologic Signature	GSE14308 TH2 VS INDUCED TREG UP
C7 Immunologic Signature	GSE14308 TH2 VS NAIVE CD4 TCELL UP
C7 Immunologic Signature	GSE14308 TH2 VS NATURAL TREG UP
C7 Immunologic Signature	GSE14308 TH2 VS TH1 UP
C7 Immunologic Signature	GSE14308 TH2 VS TH17 UP
C7 Immunologic Signature	GSE14350 TREG VS TEFF UP
C7 Immunologic Signature	GSE1460 CD4 THYMOCYTE VS NAIVE CD4 TCELL ADULT BLOOD UP
C7 Immunologic Signature	GSE1460 DP THYMOCYTE VS NAIVE CD4 TCELL ADULT BLOOD UP
C7 Immunologic Signature	GSE1460 DP VS CD4 THYMOCYTE UP
C7 Immunologic Signature	GSE1460 INTRATHYMIC T PROGENITOR VS CD4 THYMOCYTE UP
C7 Immunologic Signature	GSE1460 INTRATHYMIC T PROGENITOR VS DP THYMOCYTE UP
C7 Immunologic Signature	GSE1460 INTRATHYMIC T PROGENITOR VS NAIVE CD4 TCELL ADULT BLOOD UP
C7 Immunologic Signature	GSE15659 NAIVE CD4 TCELL VS ACTIVATED TREG UP
C7 Immunologic Signature	GSE15659 NAIVE CD4 TCELL VS RESTING TREG UP
C7 Immunologic Signature	GSE15659 RESTING VS ACTIVATED TREG UP
C7 Immunologic Signature	GSE15659 TREG VS TCONV UP
C7 Immunologic Signature	GSE17580 TREG VS TEFF UP
C7 Immunologic Signature	GSE21379 TFH VS NON TFH CD4 TCELL UP
C7 Immunologic Signature	GSE21380 NON TFH VS TFH CD4 TCELL UP
C7 Immunologic Signature	GSE22601 CD4 SINGLE POSITIVE VS CD8 SINGLE POSITIVE THYMOCYTE UP
C7 Immunologic Signature	GSE22601 DOUBLE NEGATIVE VS CD4 SINGLE POSITIVE THYMOCYTE UP
C7 Immunologic Signature	GSE22601 DOUBLE NEGATIVE VS CD8 SINGLE POSITIVE THYMOCYTE UP
C7 Immunologic Signature	GSE22601 DOUBLE NEGATIVE VS DOUBLE POSITIVE THYMOCYTE UP

C7 Immunologic Signature	GSE22601 DOUBLE NEGATIVE VS IMMATURE CD4 SP THYMOCYTE UP
C7 Immunologic Signature	GSE22601 DOUBLE POSITIVE VS CD4 SINGLE POSITIVE THYMOCYTE UP
C7 Immunologic Signature	GSE22601 DOUBLE POSITIVE VS CD8 SINGLE POSITIVE THYMOCYTE UP
C7 Immunologic Signature	GSE22601 IMMATURE CD4 SINGLE POSITIVE VS CD4 SINGLE POSITIVE THYMOCYTE UP
C7 Immunologic Signature	GSE22601 IMMATURE CD4 SINGLE POSITIVE VS CD8 SINGLE POSITIVE THYMOCYTE UP
C7 Immunologic Signature	GSE22601 IMMATURE CD4 SINGLE POSITIVE VS DOUBLE POSITIVE THYMOCYTE UP
C7 Immunologic Signature	GSE22886 CD8 VS CD4 NAIVE TCELL UP
C7 Immunologic Signature	GSE22886 NAIVE CD4 TCELL VS 12H ACT TH1 UP
C7 Immunologic Signature	GSE22886 NAIVE CD4 TCELL VS 12H ACT TH2 UP
C7 Immunologic Signature	GSE22886 NAIVE CD4 TCELL VS 48H ACT TH1 UP
C7 Immunologic Signature	GSE22886 NAIVE CD4 TCELL VS 48H ACT TH2 UP
C7 Immunologic Signature	GSE22886 NAIVE CD4 TCELL VS MEMORY TCELL UP
C7 Immunologic Signature	GSE22886 NAIVE CD4 TCELL VS NKCELL UP
C7 Immunologic Signature	GSE22886 NAIVE CD8 TCELL VS MEMORY TCELL UP
C7 Immunologic Signature	GSE22886 NAIVE CD8 TCELL VS NKCELL UP
C7 Immunologic Signature	GSE22886 NAIVE TCELL VS NKCELL UP
C7 Immunologic Signature	GSE22886 NAIVE VS MEMORY TCELL UP
C7 Immunologic Signature	GSE22886 TH1 VS TH2 12H ACT UP
C7 Immunologic Signature	GSE22886 TH1 VS TH2 48H ACT UP
C7 Immunologic Signature	GSE24574 BCL6 HIGH TFH VS NAIVE CD4 TCELL UP
C7 Immunologic Signature	GSE24574 BCL6 HIGH TFH VS TCONV CD4 TCELL UP
C7 Immunologic Signature	GSE24574 NAIVE VS TCONV CD4 TCELL UP
C7 Immunologic Signature	GSE32901 NAIVE VS TH1 CD4 TCELL UP
C7 Immunologic Signature	GSE7852 LN VS THYMUS TCONV UP
C7 Immunologic Signature	GSE7852 LN VS THYMUS TREG UP
C7 Immunologic Signature	GSE7852 TREG VS TCONV UP

C2 Curated	LEE DIFFERENTIATING T LYMPHOCYTE
C2 Curated	LEE DOUBLE POLAR THYMOCYTE
C2 Curated	LEE EARLY T LYMPHOCYTE UP
C2 Curated	LEE INTRATHYMIC T PROGENITOR
C2 Curated	LEE NAIVE T LYMPHOCYTE
C2 Curated	LEE RECENT THYMIC EMIGRANT
C2 Curated	LEE SP4 THYMOCYTE
C2 Curated	WP TH17 CELL DIFFERENTIATION PATHWAY

Supplemental Table 13: The full list of markers evaluated for annotation of single-cell clusters and results of differential expression, gene set enrichment, and module scoring analyses for the purposes of single-cell cluster annotation are available through RPubs.

Analysis	Source code and results
Annotation of thymus single-cell clusters by differential expression of key markers	https://rpubs.com/eileenowens/thymusAnnot1
Annotation of thymus single-cell clusters by GSEA using gene signatures derived from a human single-cell thymus atlas	https://rpubs.com/eileenowens/ThymAnnot2
Annotation of thymus single-cell clusters by module scoring using gene signatures derived from a human single-cell thymus atlas	https://rpubs.com/eileenowens/ThymAnnot3
Annotation of lymph node single-cell clusters by differential expression of key markers	https://rpubs.com/eileenowens/LNAnnot1
Annotation of lymph node single-cell clusters by GSEA using gene signatures derived from a human immune cell atlas and atlas of circulating canine leukocytes	https://rpubs.com/eileenowens/LNAnnot2
Annotation of lymph node single-cell clusters by module scoring for gene signatures derived from a human immune cell atlas and atlas of circulating canine leukocytes	https://rpubs.com/eileenowens/LNAnnot3

Supplemental Table 14: Quality parameters for single-cell RNA-seq data of canine lymph node and thymus.
Avg = average, UMIs = unique molecular identifiers, mito = mitochondrial.

		Pre-filtering				Post-filtering			
Sample	Location	# of cells	Avg UMIs per cell	Avg genes per cell	Avg mito genes	# of cells	Avg UMIs per cell	Avg genes per cell	Avg mito genes
L165597	Lymph node	5844	7713	2164	7%	5576	7676	5576	6%
LN154803	Lymph node	11277	5512	1846	6%	11039	5424	1850	5%
LN157849	Lymph node	13126	4705	1656	5%	12823	4732	1676	4%
T154802	Thymus	9130	6098	1943	4%	8885	5840	1926	4%
T165635	Thymus	6480	7902	2129	5%	6264	7936	2164	4%

Th157850	Thymus	7836	3934	1478	6%	7505	4020	1509	5%

Supplemental Table 15: Moran's *I* and specificity values for variable genes across pseudotime discussed in Chapter Two. A Moran's *I* value of 0 indicates no effect of a cell's position in UMAP space on gene expression, while a value of 1 indicates perfect positive autocorrelation. "Specificity" is a measurement of how specific the gene's expression is to the cell group based on Jensen-Shannon divergence and ranges from 0 to 1. "Pseudo R-squared" is a measure of how well the gene expression model fits the categorical data relative to the null model and ranges from 0 to 1.

Gene	Moran's I	p value	q value	Cluster	Fraction expressing	Specificity	Pseudo R-squared
NEDD4L	0.71	0	0	DP Thymocytes 2	0.91	0.28	0.23
CD1C	0.69	0	0	DP Thymocytes 1	0.95	0.37	0.50
NPR3	0.67	0	0	DP Thymocytes 1	0.90	0.31	0.38
INPP4B	0.67	0	0	SP Thym	0.84	0.20	0.06
DHDDS	0.68	0	0	Proliferating DP Thym	0.96	0.45	0.48
GIMAP4	0.59	0	0	CD8 NK	0.94	0.33	0.41
DNTT	0.63	0	0	DP Thymocytes 1	0.89	0.31	0.33
PAK3	0.29	0	0	ETP and DN Thym	0.78	0.36	0.29
FBN2	0.39	0	0	ETP and DN Thym	0.82	0.42	0.37
DPP4	0.35	0	0	ETP and DN Thym	0.79	0.48	0.38
TBC1D9	0.26	4.00E-12	6.54E-12	ETP and DN Thym	0.71	0.45	0.29
MRC1	0.50	0	0	ETP and DN Thym	0.69	0.68	0.37
BAHCC1	0.26	0	0	ETP and DN Thym	0.68	0.44	0.30
ALPL	0.54	0	0	ETP and DN Thym	0.58	0.78	0.32
PTPRF	0.24	0	0	ETP and DN Thym	0.53	0.53	0.23
FAU	0.73	0	0	Naive T 2	1.00	0.20	0.52
FAU	0.73	0	0	Naive to Activated T	1.00	0.19	0.43
RPL13A	0.72	0	0	Naive T 1	1.00	0.18	0.36
RPL13A	0.72	0	0	Naive T 2	1.00	0.22	0.58
RPL13A	0.72	0	0	Naive to Activated T	1.00	0.19	0.45
RPL19	0.72	0	0	Naive T 2	1.00	0.21	0.55
RPS11	0.71	0	0	Naive to Activated T	1.00	0.19	0.41
RPS15A	0.71	0	0	Naive T 1	1.00	0.17	0.33
RPS15A	0.71	0	0	Naive T 2	1.00	0.21	0.53
RPS15A	0.71	0	0	Naive to Activated T	1.00	0.19	0.45
SELL	0.63	0	0	Naive T 1	0.97	0.22	0.33
MAF	0.63	0	0	Activated T	0.76	0.56	0.38
PARP8	0.37	0	0	CD8 NK	0.93	0.30	0.35
CBLB	0.47	0	0	CD8 NK	0.95	0.34	0.41
LTB	0.73	0	0	Naive T 1	1.00	0.21	0.38
LTB	0.73	0	0	Naive T 2	0.99	0.24	0.43
LTB	0.73	0	0	Naive to Activated T	0.99	0.23	0.38

B2M	0.68	0	0	Activated T	1.00	0.19	0.40
SCML4	0.63	0	0	Naive T 1	0.97	0.21	0.31
SCML4	0.63	0	0	SP Thym and Naive T	0.67	0.26	0.05
DLA-DRA	0.75	0	0	Naive T 1	0.96	0.20	0.28
DLA-DRA	0.75	0	0	Naive T 2	0.96	0.22	0.32
DLA-DRA	0.75	0	0	Naive to Activated T	0.96	0.20	0.28
HLA-DQB2	0.64	0	0	Activated T	0.89	0.22	0.21
DLA-64	0.77	0	0	Activated T	1.00	0.21	0.39
FSTL4	0.69	0	0	DP Thymocytes 1	0.90	0.28	0.34
FSTL4	0.69	0	0	DP Thymocytes 2	0.74	0.31	0.18
TOX2	0.52	0	0	SP Thym	0.71	0.33	0.22

Appendix 3: Supplementary material for Chapter Three

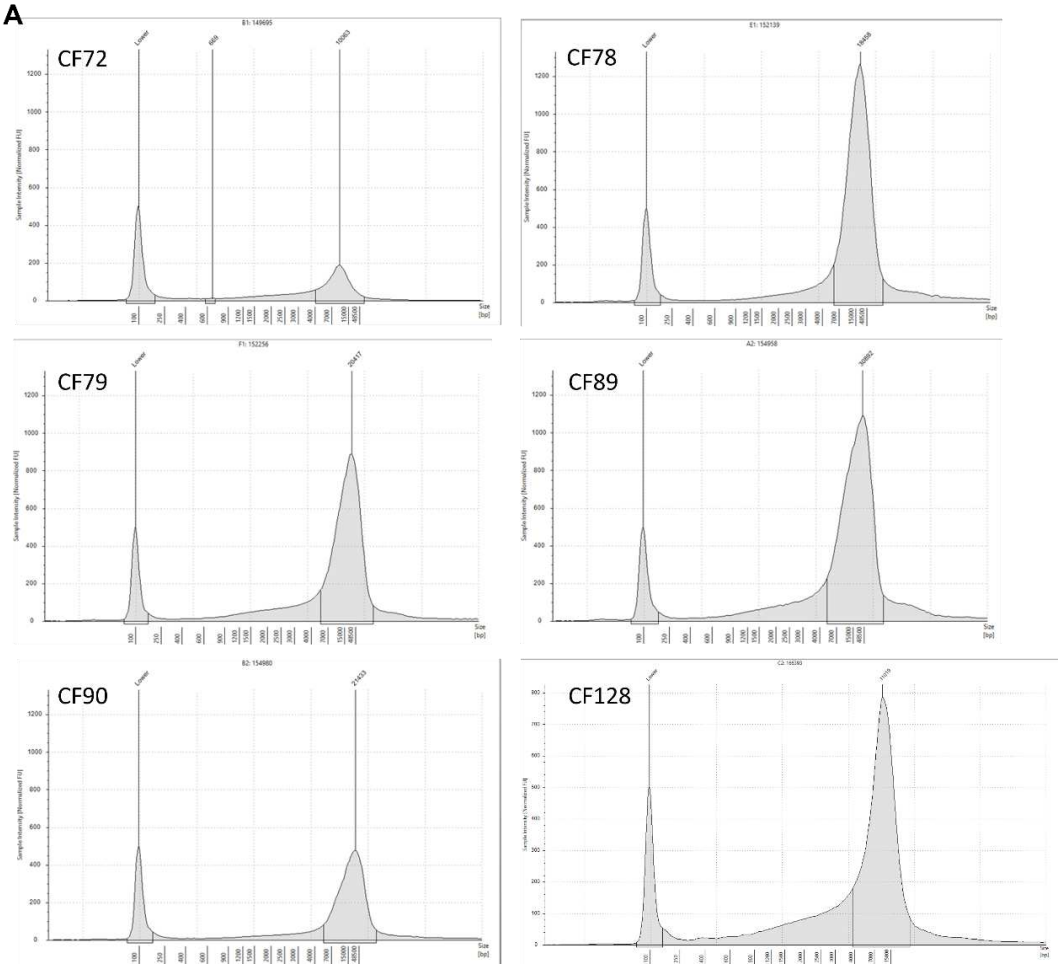
Data availability

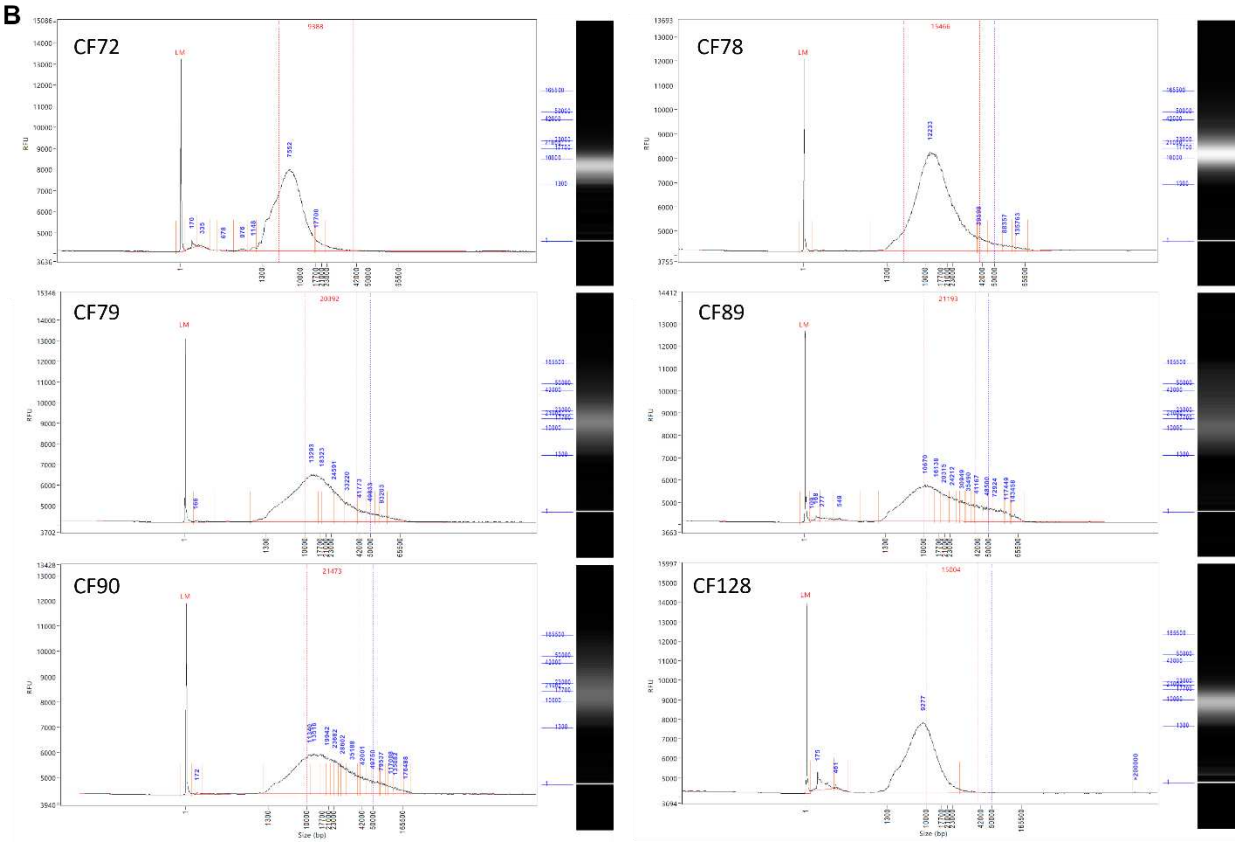
Scripts and software packages used for processing and aligning HiFi reads, performing structural variant calling, and annotating structural variant calls are documented in a Github repository: https://github.com/pirateyoho/Canine-PTCL-Long-Read-Seq-Variant-Analysis.

Source code with results for Chapter 3 figures is also available on RPubs:

Figure	Source code
Figure 62	https://rpubs.com/eileenowens/sv-analysis

A





Supplemental Figure 10: Tapestation (**A**) and Femto Pulse (**B**) tracings of the distribution of input CD4$^+$ PTCL DNA fragment sizes for long-read sequencing.

Supplemental Table 16: Sample information for CD4$^+$ PTCLs screened for fusion gene candidates by PCR. M = male. MC = male castrated. F = female. FS = female spayed. U = unknown.

Sample ID	Sample location	Breed	Sex	Age (yrs)
CF49	Lymph node	Mixed breed	MC	8
CF50	Mediastinum	Papillon	MC	11
CF51	Lymph node	Australian Shepherd	FS	5
CF52	Lymph node	English Cocker Spaniel	FS	6
CF53	Lymph node	Boykin Spaniel	M	6
CF54	Lymph node	Labrador retriever	M	4
CF55	Lymph node	Mixed breed	MC	2
CF56	Lymph node	Yorkshire Terrier	MC	7
CF57	Lymph node	Mixed breed	MC	12
CF58	Lymph node	Golden Retriever	MC	7
CF59	Lymph node	Mixed breed	FS	5
CF60	Lymph node	Golden Retriever	MC	5
CF61	Lymph node	Mixed breed	MC	4
CF62	Lymph node	Boxer	MC	7

CF63	Lymph node	Australian Shepherd	MC	5
CF64	Lymph node	Saint Bernard	MC	9
CF65	Lymph node	Border Collie	MC	7
CF66	Lymph node	Boxer	FS	4
CF67	Lymph node	Mixed breed	MC	6
CF68	Lymph node	Mixed breed	MC	6
CF69	Lymph node	Boxer	MC	7
CF70	Lymph node	Boxer	F	4
CF71	Lymph node	Boxer	FS	7
CF72	Lymph node	Old English Sheepdog	FS	4
CF73	Lymph node	Labradoodle	MC	5
CF74	Lymph node	German Shepherd	MC	7
CF75	Lymph node	Mixed breed	FS	6
CF76	Lymph node	Pit Bull Terrier	MC	5
CF77	Lymph node	Mixed breed	MC	4
CF78	Lymph node	Springer Spaniel	MC	6
CF79	Lymph node	Siberian Husky	MC	4
CF80	Lymph node	Australian Shepherd	MC	9
CF81	Lymph node-mesenteric	Mixed breed	FS	5
CF82	Lymph node	Bullmastiff	F	4
CF83	Lymph node	Boxer	MC	4
CF84	Lymph node	Mixed breed	MC	6
CF85	Lymph node	Welsh Corgi	FS	8
CF86	Lymph node	Golden Doodle	FS	6
CF87	Lymph node	American Staffordshire terrier	MC	3
CF88	Lymph node	Boxer	FS	5
CF89	Lymph node	Mixed breed	MC	8
CF90	Lymph node	Shih Tzu	MC	9
CF91	Lymph node	English Setter	F	5
CF92	Lymph node	Mixed breed	MC	6
CF93	Lymph node	Boxer	FS	8
CF94	Lymph node-mesenteric	Catahoula	MC	7
CF95	Mediastinum	Golden Retriever	M	5
CF96	Lymph node	Airedale Terrier	MC	10
CF97	Lymph node	Mixed breed	MC	6
CF98	Lymph node	Cane Corso	MC	10
CF99	Lymph node	Boxer	MC	7
CF100	Lymph node	Mixed breed	MC	6
CF101	Lymph node	Labrador retriever	FS	10
CF102	Lymph node	Golden Doodle	MC	4
CF103	Lymph node	Boxer	FS	10

CF104	Lymph node	Weimaraner	FS	5
CF105	Lymph node	Boxer	MC	5
CF106	Lymph node	German Shepherd	FS	10
CF107	Lymph node	Mixed breed	MC	7
CF108	Lymph node	Boxer	MC	10
CF109	Lymph node	Australian Cattle Dog	MC	4
CF110	Lymph node	Boxer	FS	5
CF111	Lymph node	Yorkshire Terrier	FS	9
CF112	Lymph node-mediastinal	Welsh Corgi	MC	5
CF113	Lymph node	Boxer	FS	4
CF114	Lymph node	Mixed breed	FS	11
CF115	Lymph node	Australian Shepherd	FS	5
CF116	Lymph node	Mixed breed	FS	5
CF117	Lymph node	Mixed breed	MC	8
CF118	Lymph node	Mixed breed	FS	3
CF119	Lymph node	Boxer	MC	8
CF120	Lymph node	Boxer	FS	7
CF121	Lymph node	Boxer	FS	5
CF122	Lymph node	Golden Retriever	FS	6
CF123	Lymph node	Mixed breed	MC	8
CF124	Lymph node	Blue Heeler	FS	4
CF125	Lymph node	Pit Bull Terrier	FS	3
CF126	Lymph node	American Bulldog	FS	5
CF127	Lymph node	Golden Retriever	MC	8
CF128	Lymph node	Boxer	MC	7
CF129	Lymph node	Boxer	MC	7
CF130	Lymph node	Australian Shepherd	MC	12
CF131	Lymph node	Mixed breed	MC	13
CF132	Lymph node	Mixed breed	MC	8
CF133	Lymph node	Boxer	FS	6
CF134	Lymph node	Old English Sheepdog	MC	3
CF135	Lymph node	Mixed breed	MC	4
CF136	Lymph node	Mixed breed	MC	5
CF137	Lymph node	Boxer	MC	7
CF138	Lymph node	Golden Doodle	FS	8
CF139	Lymph node	Mixed breed	FS	6
CF140	Lymph node	Boxer	MC	9
CF141	Lymph node	Boxer	MC	11
CF142	Lymph node	Newfoundland	MC	4
CF143	Lymph node	Mixed breed	M	5
CF144	Lymph node	Golden Doodle	MC	7

CF152	Peripheral blood	Border Collie	MC	11
CF153	Lymph node	Boxer	MC	5
CF154	Lymph node	Mixed breed	MC	4
CF155	Lymph node	Mixed breed	MC	7
CF156	Lymph node	WGRIF	FS	7
CF157	Lymph node	Pit Bull Terrier	M	4
CF158	Lymph node	Golden Retriever	MC	7
CF159	Lymph node	Australian Shepherd	MC	5
CF160	Lymph node	Boxer	FS	9
CF161	Lymph node	Shetland Sheepdog	FS	8
CF162	Lymph node	American Bulldog	MC	4
CF163	Lymph node	Collie	MC	6
CF164	Lymph node	Golden Retriever	FS	6
CF165	Lymph node	Golden Retriever	MC	10
CF166	Lymph node	Boxer	F	6
CF167	Lymph node	Mixed breed	FS	6
CF168	Lymph node	Golden Retriever	MC	4
CF169	Lymph node	Mixed breed	FS	5
CF170	Lymph node	Sealyham Terrier	FS	3
CF171	Lymph node	Golden Retriever	U	3
CF172	Mediastinum	Golden Retriever	FS	9
CF173	Lymph node	Pit Bull Terrier	MC	8
CF174	Lymph node	Mixed breed	MC	10
CF175	Lymph node	Labradoodle	F	6

Supplemental Table 17: Sample information for $CD4^+$ PTCLs screened for fusion gene candidates by PCR.

Sample ID	**Sample location**	**Breed**	**Sex**	**Age (yrs)**	**Reported concurrent conditions**	**Lymphoma ruled out by flow cytometry**
CF145	Lymph node sorted CD4	Mixed breed	F	1	None	Yes
CF146	Lymph node sorted CD4	Mixed breed	F	1	None	Yes
CF147	Lymph node sorted CD4	Mixed breed	F	1	None	Yes
CF148	Thymus sorted CD4	Mixed breed	F	0.66	None	Yes
CF149	Lymph node sorted CD4	Mixed breed	F	0.66	None	Yes

CF150	Lymph node sorted CD4	Mixed breed	F	0.73	None	No
CF151	Thymus sorted CD4	Mixed breed	F	0.73	None	No
CF176	Whole blood	Mixed breed	MC	10	Immune-mediated hemolytic anemia	No
CF177	Whole blood	Boxer	MC	8	Acanthomatous ameloblastoma	No
CF178	Whole blood	Mixed breed	FS	10	Immune-mediated thrombocytopenia	No
CF179	Whole blood	Mixed breed	FS	9	Immune-mediated hemolytic anemia	No
CF180	Whole blood	Mixed breed	FS	12	MCT, nasal carcinoma	No
CF181	Whole blood	Labrador Retriever	MC	7	None	No
CF182	Whole blood	Mixed breed	FS	9	None	No
CF183	Blood	Golden Retriever	FS	2	None	Yes
CF184	Blood	German Shorthaired Pointer	FS	9	None	Yes
CF185	Blood	Labrador Retriever	MC	3	None	Yes

Supplemental Table 18: Sample details for CD4⁺ PTCLs screened for structural variants by long-read sequencing. LN = lymph node. OESD = Old English Sheepdog. SPRSP = Springer Spaniel. HUS = Siberian Husky. MIX = mixed breed dog. SHTZ = Shih Tzu. BOX = Boxer. FS = female spayed. MC = male castrated.

ID	**Site**	**Breed**	**Sex**	**Age (yrs)**	**Conc. (ng/ uL)**	**Total (ug)**	**Bar-code quality**	**HiFi Reads**	**Mean HiFi Read Length (bp)**	**Median HiFi Read Quality**	**Fusions + by PCR**
CF72	LN	OESD	FS	4	285	4.8	97.1	3,735,101	7,507	Q37	None

CF78	LN	SPRSP	MC	6	1000	17	96.9	3,474,221	6,660	Q38	*JPT1-GATD3A, PTMA-HMGB1, PTMA-LMO4, REV3L-FYN*
CF79	LN	HUS	MC	4	580	9.8	96.9	5,733,070	5,614	Q40	*JPT1-GATD3A, NCL-PTMA, SRSF5-LMO4, TPD52L2-GATD3A, PER1-EIF5A, MYC-TRIB1, PTMA-LMO4, REV3L-FYN*
CF89	LN	MIX	MC	8	230	3.9	97.4	3,766,001	6,639	Q39	*JPT1-GATD3A, NCL-PTMA, PER1-EIF5A, PTMA-HMGB1, REV3L-FYN*
CF90	LN	SHTZ	MC	9	360	6.1	96.9	3,833,417	6,989	Q37	*B2M-GNAS, JPT1-GATD3A, MROH1-GATD3A, NCL-PTMA, SRSF5-LMO4, TPD52L2-GATD3A, PTMA-LMO4, REV3L-FYN*
CF128	LN	BOX	MC	7	56	0.95	97	3,715,223	7,553	Q37	*MROH1-GATD3A, SRSF5-LMO4, TPD52L2-GATD3A, MYC-TRIB1, PTMA-LMO4, REV3L-FYN, KLF10-YWHAZ*

Supplemental Table 19: Sequences of fusion gene candidates selected for validation by PCR from FusionCatcher and STAR-Fusion outputs. Sequences are based on the CanFam3.1 reference genome, with the exception of *CHD3-EIF5A*,

which is based on the ROS Cfam 1.0 reference genome. The sequences associated with the 3 different predicted breakpoints for the *GRB10-IKZF1* fusion are represented on different rows.

REV3L-FYN
CTGCCGCCGCCAGATCGCCCGTGAGGAGGAGGAGGAGGAGGCGGTGGCGGCGGCGGCGAACATGTTTTCAGTGA GGATAGTGACCGCCGACTACTACATGGCCAGCCCGCTGCGGGGCCTGGACACCTGCCTGTGCCCCCTCACCCAG GCCCCTGTCAAGAAGGTGCCGGTGGTGCGAGTCTTCGGAGCGACCCCGGCAG*GCATGAACAACCGCGAGGTGC TGGAGCAGGTGGAGCGCGGCTACAGGATGCCCTGCCCGCAGGACTGCCCCATCTCCCTCCACGAGCTCATGATC CACTGCTGGAAGAAGGACCCCGAGGAACGGCCCACCTTCGAGTACTTGCAGGGCTTCCTGGAAGACTACTTCAC GGCCACGGAGCCCCAGTACCAGCCCGGCGAG
NCL-PTMA
CACTCCGCCGGCATAAGCCGCCATCATGGTGAAGCTCGCAAAG*GACTTAAAGGAGAAGAAGGAAGTTGTGGAG GAGGCAGAGAATG
PTMA-LMO4
TTTGCATTGTTCCTGTGTCCGGCTCCCCGTTCGCCGCAACCACCTCCGCCGCGCGCCTCCTCCGCCTCCGCGGAC TCCGGCAGCTCCCTCGCCAGAGTCCTCGAACTCTTGCTCTCTTTTTAATCCGCTGCATCGGATCACCGGCGTGCC CCATCATGTCAGACGCGGCCGTGGACACCAGCTCCGAGATCACCACCAAG*ACCATGGTGAATCCGGGCAGCAGC TCACAGCCGCCCCCCGTGACGGCCGGCTCCCTCTCCTGGAAGCGGTGTGCAGGCTGCGGAGGCAAGATCGCGGA CCGCTTTCTGCTCTATGCCATGGACAGCTACTGGCACAGCCGGTGCCTCAAGTGCTCCTGCTGCCAGGCGCAGCT GGGCGACATCGGCACGTCCTGTTACAC
SRSF5-LMO4
ACGGACCTCTCTGGTCTCAGCTGCCAAAGATCCTGTCCGGTAG*ACCATGGTGAATCCGGGCAGCAGCTCACAGC CGCCCCCCGTGA
TPD52L2-GATD3A
CCACCGTGGGCTCCGCCATTAGCAGGAAGCTCGGGGATATGAG*CTTGTGCTGCATCGCGTCTGTCCTCGCAGCC AAAGTGCTCCAT
MYC-TRIB1
GCTCCCCCCCTCCTCCCCTTTTTGGAACTTACAACCACCCGAGAGCGAGGAGAACGCGACTCTCCGGACGCGCAG AGGCTCTACCCCTTCGCCTATTTGGGAAGACACTTCTCCCCGCCGCCGCCCGCGCCCGGCTGCCCTGAAAGGCGC TCCCCGCCGCATTTTGGACGCTGGATCTCCTCCGGAGAGTGGAAACCCCG*AAGCCAGCTCAGACTGGAAAGTCT AGAAGACACACACATAATCAAGGGGGAAGACGATGCTCTGTCAGACAAGCACGGCTGCCCGGCCTACGTGAGCC CTGAGATTCTGAACACCACGGGGACCTACTCCGGAAAGGCGGCAGACGTTTGGAGCCTAGGGGTGATGCTCTAC ACCCTTTTGGTGGGACGATACCCCTTCC
B2M-GNAS
GGCCTTGCTCCTCATCCTCCTCGCTGCGTGCCGCCTGGATGCCGTGCAGC*GTGCCGGAGAATCTGGTAAAAGCA CCATTGTGAAGCAGATGAGGATCCTG
PTMA-HMGB1
TTTGCATTGTTCCTGTGTCCGGCTCCCCGTTCGCCGCAACCACCTCCGCCGCGCGCCTCCTCCGCCTCCGCGGAC TCCGGCAGCTCCCTCGCCAGAGTCCTCGAACTCTTGCTCTCTTTTTAATCCGCTGCATCGGATCACCGGCGTGCC CCATCATGTCAGACGCGGCCGTGGACACCAGCTCCGAGATCACCACCAAG*AAAATAATTAAACATGGGCAAAGG AGATCCTAAGAAGCCGAGAGGCAAAATGTCATCATATGCATTCTTTGTGCAAACTTGCCGAGAGGAGCACAAGAA GAAGCACCCAGATGCTTCAGTCAACTTCTCAGAGTTTTCTAAGAAGTGCTCAGAAAGGTGGAAGGTAAGGGGGC TTGAAAAATATTAAGGAGGTGATTTTT
JPT1-GATD3A
CCACCTTCAAGGGAGTCGATCCCAACAGCAGGAATAGCTCTCG*CTTGTGCTGCATCGCGTCTGTCCTCGCAGCC AAAGTGCTCCAT
KLF10-YWHAZ
GCGCGGGGCAGGCGGAGAGGCTACCTGAATGGGGAGGGGGCAGACGGCGCGGCGCGCGGCGGCGGCGGGAGCG GCGTCGAGTGTCTCCGAGCGCCCGTCCGTGGCCCGGCAGCCAAGCAGCGCAGCAGCCAGTGAGCTCGCCGCCGG CCGCCAGGCAGCCAGCCATGCTCAACTTCGGCGCTTCCCTCCAGCTCGCTGCG*AGCATCCAGTCATGGATAAAA ATGAGCTGGTGCAGAAGGCCAAACTGGCCGAGCAGGCTGAGCGATATGATGACATGGCAGCCTGCATGAAGTCC GTAACTGAGCAAGGAGCTGAATTATCCAATGAGGAGAGGAATCTTCTCTCAGTTGCTTATAAAAATGTTGTAGGA GCCCGTAGGTCATCTTGGAGGGTCGTCTCA
MROH1-GATD3A

CGGTGGTCACCAGCCTCCTGGGAAGCCCCCTGCCCTTTGACAG*CTTGTGCTGCATCGCGTCTGTCCTCGCAGCC AAAGTGCTCCAT
TOX2-LMO4
GCCGGCCGGCCTGGCGCACCTGGACTATTACCACTGCGGCAAG*ACCATGGTGAATCCGGGCAGCAGCTCACAGC CGCCCCCCGTGA
PER1-EIF5A
TCGGGGGCGGGGCCTGCGGCTCCGCGGCAGGAGGCCAATGTGGGGCAGTGCCTGGCATTATGCAACCCGCCTCC TGGCCCCGCGAAGCTTCCACTCGGCGGAGGGCTGCAACGGCGGCGGGCAGGAGGCCGGAGAGAGCTCGCGCGGC CAG*TTGGAATCGAAGTCTCTTAAAATGGCAGATGATTTGGACTTCGAGACAGGAGATGCAGGGGCCTCAGCCAC CTTCCCAATGCAGTGCTCAGCATTACGTAAGAACGGCTTCGTGGTGCTTAAAGGCCGGCCATGTAAGATCGTGGA GATGTCTACCTCCAAGACTGGCAAGCACGGCCATGCCAAGGTTAGAACTTGACC
BZW1-HSPD1
GAGGCGGAGGCTCGAGGCCGCGACCGCGAGGGACGCGGGGGTGGACGCCCCGAGGAAGAAATCTCGCGATAGG AGGCTGGCCGGGGCGGGGGTGACACCTCTTCCCTCCTCCTCGCGTTAGTTCCGGTCGCAGAGGAGACACCGCCG CAGTTGCCGCCACATCGGGGATTTCTGGCTCTTTTCTCTTCGCCTTAAATTCG*AAATGCTTCGATTACCCGCAGT GCTCCGCCACGTCAGGCCAGTGTCCAGAGCCCTGGCTCCTCATCTCACTCGGGCCTATGCCAAGGATGTAAAATT CGGTGCCGATGCCCGAGCCTTAATGCTTCAAGGTGTAGATCTCCTAGCCGATGCCGTAGCCGTTACTATGGGGCC AAAGGTATCAGTGTTTTTTGTTTTGTTG
CHD3-EIF5A
GGAAGAAGAGGGCGACGAGGAGGAGGAGGAGGTGGAGGCGGCCGACGAGGACTACGAGGAGGACGACGACGAG GGAGTGCTCGGGCGCGGGCCGGGCCACGACCGGGGCCGCGACCGCCACAGCCCCCCCGGCTGCCACCTCTTCCC GCCGCCGCCGCCGCCGCCGCTGCCCCCGCCGCCGCCGCCGCCCCCGCCGCCAG*TTGGAATCGAAGTCTCTTAAA ATGGCAGATGATTTGGACTTCGAGACAGGAGATGCAGGGGCCTCAGCCACCTTCCCAATGCAGTGCTCAGCATTA CGTAAGAACGGCTTCGTGGTGCTTAAAGGCCGGCCATGTAAGATCGTGGAGATGTCTACCTCCAAGACTGGCAA GCACGGCCATGCCAAGGTTAGAACTTGACC
BCL11B-MYH9
TGTGCAACTACGCGTGCGCGCAGAGCAGCAAGCTCACGCGCCACATGAAG*GCGCCACTGTCCCCTCCTGTTTCT AAAGTGTCTCAAGTGCAATGACGACC
DENND4B-MYH9
TCCTTCCCGCCGCTCCCCCGCCTCTCATGTTCCCCCACCTGAGCTGCCCC*CCAGTTGCTGTCACAAAGGAGTGA ACGGAGCTCCCAGCCCCACCCCTTTT
GRB10-IKZF1
CCAGGCTTCCCGGCGCCGGCCGACCGGCTCACAGGCCACCAGG*ATAACCTGAGGACAATGGATGCTGATGAGG GTCAAGACATGTC
TATTACATATTTTTTTACGACGGCAGCTTTTAACCCGCGGTTGTGTCTCTCCCCCCAGACGCCTGCAGGAAGAAG ACCAGCAGCTCAGAACCTCGTCTCTGCCGGCCATCCCCAACCCCTTCCCCGAGCTCTGCGGCCCCGGGAGCCCGC CGGTGCTGACGCGCGGCTCCCTACCTCCAGGCCAGGCAGCTGCCAAGCAG*ATAACCTGAGGACAATGGATGCTG ATGAGGGTCAAGACATGTCCCGAGTTTCAGGTGAGACCTTCGTGGGGTACCTGAGTGGGGGGCCCGTGAGGAGC CATGTCTGGGGTGCCGGGAGCCCAGGCGCTCGCTCCGTTCCCGAGGGAGGGAGATAGGCCGATCCGCAGAAGGA GGGACCCGAGACGGAATCGGATGCTTTA
CGGCCTGAGCCCTGTCCAGGGGCTGCCCGCTGGGACGGGTCGCTTGCAGTCAAGTGGTGAGCAGCGAACTCGTC CTCTGGTCTGGCTCTGACGGTCTGTCTTTGTTGCAGGTGCTGAGGAGCGGTAAATGTAGCTCGAAGAAGGAACC CATGGCTTTGGCTGGCTGCCCAGATTCCTTCTTGCACCACCCGTACTACCAG*CCAGTAATGTTAAAGTAGAGGC TCAGAGTGATGAAGAGAATGCGCGTGCCTGTGAGCTGAATGGGGAAGAATGTGCGGAGGATTTACGAGTGCTTG ATGCTGCGGGAGAGAAAATGAATGGCTCCCACGGTGGCCAAGGCGGCAGAGCTTTGTCGGGAGCTGGAGGCATT CGACTTCCTAACGGCAAACTAAAGTGTGAT
ITM2C-XPO1
GGCCCCGGCTCCGGCCGCTGAGATCCTGCTGACGCCGGCTCGG*TAATCTATGCCAGCAATTATGACAATGTTAG CAGACCATGCAG
MCM5-PTMA
GCTTTACCAGCCAGGAGGACCAGGAGCTTCTGAGCCGCATCGAGAAGCAG*GACGTCGACACCAAGAAGCAGAA GACTGATGAGGATGACTAGACAGCAAA

NFKB2-TRIM8
CCCCCTGAGCCACCAGGAGGGCTCTGCCATGGGCACCCCCAGCCTCAG*AACACCAAGTCTGTGAAAATCCTAATGGACAGGTGAGCAGATGACCACTG
PCGF3-IKZF1
AACACTTCACAGTCTCAGCGGCCGGAGAGGCGCCTCTCGCAGACTAGATG*GGGAGCGGCCCTTCCAGTGCAACCAGTGCGGCGCCTCCTTCACGCAGAAG
RGS10-PTMA
ACAAGTCAACGTCGAGGGGCAGTCCCGGCTCAATGAGAAGATACTGGAAG*CCGCGGACTCCGGCAGCTCCCTCGCCAGAGTCCTCGAACTCTTGCTCTCT
SPOCK2-XPO1
TGGTTCCAGCTCCTTCATGAGAACGCCAAGCAGAATGGCTCGGCCGGCAG*GGATATTTGTCGACCAAAATGATGCCAATTTGTAAATTAAAATGTCACCT
TFAM-MYH9
GTTATTATAATGAAATGAAATCTTGGGAAGAACAGATGATGGAAGTCGGA*GGGGACTCCACCGACCTCAATGACCAGATCGCCGAGCTGCAGGCTCAGAT
VBP1-JUN
GATACAGTGCTGAAGAAGCTGGATGAACAGTACCAGAAGTACAAGTTTAT*CCGCTGCTCAGCGCGCCCGGGTTGAAGTTGCTGAGGTTGGCGTAGACCGG
YWHAZ-KLF10
TGCTGAGCGCCCGTCCGTCCGCCGCCAGCCACTCCGGACACAG*GAGGAAAGGATGGAAATAATTTCTGAAAGATCAAAAGAGAGTG

Supplemental Table 20: Summary of PCR results for each fusion gene candidate selected for our PCR assay. Normal samples were only screened for fusion gene candidates that had been previously successfully validated by PCR in $CD4^+$ PTCL samples.

Fusion Gene	Detected by PCR in $CD4^+$ PTCL	Detected by PCR in control sorted $CD4^+$ thymocytes	Detected by PCR in control sorted nodal $CD4^+$ lymphocytes	Detected by PCR in normal dog whole blood
REV3L-FYN	Yes	Yes	Yes	Yes
NCL-PTMA	Yes	Yes	Yes	Yes
PTMA-LMO4	Yes	No	No	Yes
SRSF5-LMO4	Yes	No	No	Yes
TPD52L2-GATD3A	Yes	Yes	Yes	Yes
MYC-TRIB1	Yes	No	No	Yes
B2M-GNAS	Yes	No	Yes	Yes
PTMA-HMGB1	Yes	Yes	No	Yes
JPT1-GATD3A	Yes	No	No	Yes
KLF10-YWHAZ	Yes	No	No	Yes
MROH1-GATD3A	Yes	No	No	No
TOX2-LMO4	Yes	No	No	No
PER1-EIF5A	Yes	No	No	No
BZW1-HSPD1	No	No data	No data	No data

CHD3-EIF5A	No	No data	No data	No data
BCL11B-MYH9	No	No data	No data	No data
DENND4B-MYH9	No	No data	No data	No data
GRB10-IKZF1	No	No data	No data	No data
GRB10-IKZF1	No	No data	No data	No data
ITM2C-XPO1	No	No data	No data	No data
MCM5-PTMA	No	No data	No data	No data
NFKB2-TRIM8	No	No data	No data	No data
PCGF3-IKZF1	No	No data	No data	No data
RGS10-PTMA	No	No data	No data	No data
SPOCK2-XPO1	No	No data	No data	No data
TFAM-MYH9	No	No data	No data	No data
VBP1-JUN	No	No data	No data	No data
YWHAZ-KLF10	No	No data	No data	No data

Supplemental Table 21: Genes significantly upregulated (defined as having a log2fc ≥1 and padj < 0.05) and significantly downregulated (defined as having a log2fc < 1 and padj < 0.05) in CD4⁺ PTCLs that were PCR-positive for the *MROH1-GATD3A* fusion compared to CD4⁺ PTCLs that were PCR-negative for the *MROH1-GATD3A* fusion gene.

Upregulated in *MROH1-GATD3A*+ CD4⁺ PTCL		
Gene	**log2fc**	**padj**
ENSCAFG00000003812	3.86718616	2.19E-11
GPR85	3.549137663	0.004067
ENSCAFG00000023585	3.062642922	0.005565
GPR25	2.725596786	0.047518
EPYC	2.724115893	0.044492
ABCC8	2.700619789	8.86E-05
ENSCAFG00000043049	2.51139034	0.007734
ARPP21	2.462139621	0.04845
TGFB3	1.96629188	2.22E-05
VSNL1	1.741759141	0.029024
TFF1	1.63913482	0.041997
WLS	1.638759634	0.04679
ENSCAFG00000041361	1.416985541	0.021967
TMEM40	1.371300653	0.023535
LRP12	1.361601663	0.029565
ENSCAFG00000024259	1.356797698	0.003471
NRG2	1.254097239	0.00038
ENSCAFG00000040706	1.22744587	0.002466
ENSCAFG00000011681	1.08922796	0.04808
FAM160A1	1.002162684	0.041401

ARMH4	1.001691356	0.025675
Downregulated in *MROH1-GATD3A*+ CD4⁺ PTCL		
Gene	**log2fc**	**padj**
ADAD1	-7.082718422	0.004931
ENSCAFG00000045261	-5.159576403	0.04408
ENSCAFG00000041215	-4.935295858	0.000465
ENSCAFG00000014011	-4.903472442	6.05E-15
ENSCAFG00000029181	-4.526708184	0.00038
MUC4	-4.467328838	8.86E-05
CORIN	-4.465632942	0.000973
FGF23	-4.184010959	0.001582
TKTL1	-3.962263164	0.000109
ENSCAFG00000037290	-3.914614968	0.002447
ENSCAFG00000029953	-3.771742755	2.06E-05
ENSCAFG00000041788	-3.410824707	0.043979
ENSCAFG00000030524	-3.370908541	8.86E-05
NPR3	-3.205872801	0.001133
ABCB5	-3.081905519	0.040668
TRAV21	-3.041224352	0.000507
NAV3	-2.951529704	0.000372
ENSCAFG00000037816	-2.886089802	0.003471
IL21	-2.773511599	0.009159
COL6A5	-2.697981094	0.019263
IL4	-2.565224782	0.03434
FBP2	-2.542389378	0.000964
ENSCAFG00000046831	-2.499565485	0.000758
WNT3	-2.47005084	2.06E-05
ENSCAFG00000006763	-2.428381502	0.001438
DOK6	-2.427452585	0.000251
NRG1	-2.329579672	6.55E-12
TRBV25-1	-2.241228258	6.55E-12
PLCH2	-2.201177383	0.020915
IL2	-2.195818817	0.022692
EPHA1	-2.180828168	0.022944
IL1RL1	-2.080521086	0.010256
TBX1	-2.043923787	0.020915
ROBO1	-2.034610734	0.02173
FNDC5	-2.013713228	0.000996
RBFOX1	-1.980823652	0.004282
ENSCAFG00000042075	-1.962112002	0.010256
TENM4	-1.895677032	0.000204

SPRN	-1.836629227	0.003483
CWH43	-1.731764024	0.01105
ENSCAFG00000048980	-1.683875401	0.03434
ENSCAFG00000049585	-1.628761327	0.019384
EPHB2	-1.576670785	0.018837
ENSCAFG00000029542	-1.519406617	0.010256
HAVCR2	-1.433849855	0.004264
ENSCAFG00000046002	-1.428725746	0.002508
ENSCAFG00000028523	-1.376021017	0.001133
SLC6A9	-1.368334275	0.016016
ENSCAFG00000004676	-1.335979606	0.042928
TMEM64	-1.302677225	0.03434
ENSCAFG00000043591	-1.253277698	0.011789
ENSCAFG00000024882	-1.242458424	6.97E-05
TAL1	-1.207538111	0.03434
GCAT	-1.204151782	0.032518
TRAV29DV5	-1.158228	0.001328
APPL2	-1.12402063	0.002999

Supplemental Table 22: Genes significantly upregulated (defined as having a log2fc ≥1 and padj < 0.05) and significantly downregulated (defined as having a log2fc < 1 and padj < 0.05) in CD4$^+$ PTCLs that were PCR-positive for the *PER1-EIF5A* fusion compared to CD4$^+$ PTCLs that were PCR-negative for the *PER1-EIF5A* fusion gene.

Upregulated in *PER1-EIF5A*+ CD4$^+$ PTCL		
Gene	**log2fc**	**padj**
NPR3	5.348203	1.18E-13
ENSCAFG00000003812	4.143204	3.59E-10
IL21	4.129513	3.20E-06
ASCL2	4.950585	3.20E-06
SLC45A1	2.29763	2.36E-05
ENSCAFG00000013086	3.356078	3.38E-05
TOGARAM2	2.432695	0.000125
ENSCAFG00000033337	3.137188	0.000125
PDCD1	2.44498	0.000153
ENSCAFG00000041266	3.435803	0.00036
ANGPTL6	1.017691	0.001286
ENSCAFG00000041525	1.995241	0.00305
ARMH4	1.28083	0.005705
P3H4	1.448792	0.006792
ENSCAFG00000040706	1.338316	0.006792
L3MBTL4	1.844497	0.007544
ENSCAFG00000034152	1.048169	0.007964

NRP2	1.420939	0.008547
CXCL11	2.285093	0.008547
ENSCAFG00000046448	2.857642	0.009582
ENSCAFG00000046747	2.944066	0.011332
ENSCAFG00000029651	1.613116	0.011332
ENSCAFG00000043921	3.920286	0.016568
CD27	1.009515	0.021061
TSPAN9	1.09213	0.021511
COL4A5	2.249573	0.024545
FKBP10	1.491854	0.031506
ENSCAFG00000041446	1.352547	0.031571
DLX1	1.934606	0.032112
ENSCAFG00000049275	2.197478	0.032879
TSHZ2	1.940088	0.032882
ENSCAFG00000029652	1.634696	0.040538
AFAP1L1	1.328436	0.040538
HSPB8	1.448642	0.040538
ENSCAFG00000014952	1.938352	0.040538
PLEKHG5	1.424971	0.041133
ST8SIA2	2.473006	0.041133
ZNF697	1.882361	0.041562
ENSCAFG00000028596	1.062371	0.041793
TLN2	1.240367	0.042513
ENSCAFG00000044565	1.991552	0.049392
MCF2	3.437967	0.049392
ZFHX3	1.272781	0.049392
TNFSF4	1.576846	0.049392
Downregulated in *PER1-EIF5A*+ CD4⁺ PTCL		
Gene	**log2fc**	**padj**
TRBV24-1	-6.9236	1.30E-06
ENSCAFG00000045428	-5.3958	0.042341
CCDC172	-5.2566	0.011332
TRBV25-1	-4.04573	3.77E-11
PEG3	-3.93283	2.33E-05
ENSCAFG00000032249	-3.74172	0.005136
BICDL2	-3.65893	6.81E-06
NRG1	-3.45849	3.39E-06
CWH43	-3.33548	0.009758
TRAV21	-3.03255	0.006792
WIPF3	-2.95328	0.002655
TNFRSF19	-2.92213	0.003176

EPHA1	-2.84482	0.011713
CORIN	-2.73812	0.016568
MINAR1	-2.64733	0.046346
KCNJ2	-2.61266	0.025548
CASR	-2.60039	0.011006
DOK6	-2.56557	0.008225
ENSCAFG00000048283	-2.54965	0.039693
ALDH1A1	-2.54732	0.003633
HES7	-2.43612	0.047964
ENSCAFG00000006763	-2.42232	0.005072
ENSCAFG00000042264	-2.42148	0.035198
ENSCAFG00000041887	-2.39692	0.020941
KLRG1	-2.3353	0.001009
ENSCAFG00000043621	-2.32278	9.26E-06
SULT2B1	-2.29775	0.016396
TREM2	-2.29336	0.032112
ATP8B3	-2.27299	0.003852
ENSCAFG00000029421	-2.15301	0.007839
ENSCAFG00000032488	-2.12972	0.040538
CABLES1	-2.11497	0.000101
ENSCAFG00000037413	-2.06188	0.00132
ENSCAFG00000045674	-2.00428	0.03989
ENSCAFG00000029724	-1.95369	1.43E-06
JAG2	-1.89497	0.027518
SPTBN4	-1.88549	0.031506
ENSCAFG00000023990	-1.83221	0.049392
SH3YL1	-1.812	0.000118
STAT4	-1.78847	0.049392
PLXNA2	-1.7709	0.00757
ENSCAFG00000014011	-1.76153	0.011332
VPREB3	-1.74778	0.009368
CNTN3	-1.5742	0.040538
ENOX1	-1.40467	0.044836
HTR4	-1.40252	0.001342
CCR6	-1.39021	0.025381
CRLF2	-1.37023	0.000133
CACNA1G	-1.309	0.003082
ENSCAFG00000024337	-1.1893	0.010215
ENSCAFG00000027465	-1.18293	0.016396

Supplemental Table 23: Alignment statistics for HiFi reads using the pbmm2 aligner. Total HiFi read yield was reported in the QC report for the sequencing run generated by the Arizona Genomics Institute. All other metrics were derived using the *--log-level INFO* option during pbmm2 alignment. Gap-compressed sequence identity is calculated by pbmm2 using the following formula: 100 * #Matches / (#Matches + #Mismatches + #DeletionEvents + #InsertionEvents).

Sample	Genome	Total HiFi read yield	Mapped reads	Mapping rate	Alignments	Mapped bases	Mean Gap-Compressed Sequence Identity	Max Mapped Read Length	Mean Mapped Read Length
CF72	CanFam3.1	5,733,070	5,730,456	99.954%	5,946,666	31,933,813,427	99.5533%	45,463	5,370.04
CF72	CanFam4	5,733,070	5,731,901	99.980%	5,847,321	32,093,584,870	99.6563%	54,023	5,488.60
CF78	CanFam3.1	3,735,101	3,733,795	99.965%	3,879,040	27,858,950,687	99.5385%	61,651	7,181.92
CF78	CanFam4	3,735,101	3,734,474	99.983%	3,812,273	27,963,576,639	99.6297%	61,651	7,335.15
CF79	CanFam3.1	3,474,221	3,471,219	99.914%	3,622,078	22,935,334,169	99.4402%	66,332	6,332.09
CF79	CanFam4	3,474,221	3,472,396	99.947%	3,551,950	23,049,118,593	99.5552%	66,332	6,489.15
CF89	CanFam3.1	3,833,417	3,831,467	99.949%	3,979,941	26,599,172,719	99.5333%	70,131	6,683.31
CF89	CanFam4	3,833,417	3,832,617	99.979%	3,903,374	26,723,987,018	99.6370%	70,131	6,846.38
CF90	CanFam3.1	3,715,223	3,713,310	99.949%	3,861,127	27,851,209,844	99.5081%	70,867	7,213.23
CF90	CanFam4	3,715,223	3,714,481	99.980%	3,783,031	27,986,559,198	99.6245%	70,867	7,397.92
CF128	CanFam3.1	3,766,001	3,764,653	99.964%	3,885,721	24,852,832,460	99.6605%	40,346	6,395.94
CF128	CanFam4	3,766,001	3,765,313	99.982%	3,838,744	24,930,112,317	99.6562%	40,346	6,494.34

LIST OF ABBREVIATIONS

A - Adenine

ABL - Abelson proto-oncogene 1, non-receptor tyrosine kinase

ACKR - Atypical chemikine receptor

ADAM - A Disintegrin And Metalloprotease

AEBP - Adipocyte enhancer-binding protein

AhR - Aryl hydrocarbon receptor

AIRE - Autoimmune regulator

AITL - Angioimmunoblastic T-cell lymphoma

AKT - AKT Serine/threonine Kinase

ALCL - Anaplastic large cell lymphoma

ALK - Anaplastic lymphoma kinase

ALPL - Alkaline phosphatase, biomineralization associated

ARPP - cAMP-regulated phosphoprotein

ASAP - ArfGAP with SH3 domain, ankyrin repeat and PH domain

ASCL2 - Achaete-scute family bHLH transcription factor 2

ATAC - Assay for transposase-accessible chromatin

ATLL - Adult T-cell lymphoma/leukemia

ATM - Ataxia telangiectasia mutated

B2M - Beta-2-microglobulin

BAD - Bcl2-associated agonist of cell death

BAHCC - BAH domain and coiled-coil containing

BAM - Binary alignment map

BCAP - B cell receptor associated protein

BCL - B cell lymphoma

BCR - B cell receptor

bp - Base pairs

BRAF - B-Raf Proto-oncogene

BWA - Burrows-Wheeler Aligner

BZW - Basic leucine zipper and W2 domains

C - Cytosine

CAT - Catalase

CBLB - Cbl proto-oncogene B

CCR - C-C motif chemokine receptor

CCS - Consensus circular sequencing

CD - Cluster of differentiation

CDH - Cadherin

CDKN - Cyclin-dependent kinase inhibitor

cDNA - Complementary DNA

CDS - Coding sequence

CE - Capillary electrophoresis

CELF - CUGBP elav-like family member

CHD - Chromodomain helicase DNA binding protein

CHOP - Cyclophosphamide, doxorubicin, vincristine, prednisone

COSMIC - Catalogue of Somatic Mutations in Cancer

COX - Cytochrome C oxidase subunit

CTAT - Cancer Transcriptome Analysis Toolkit

CTLA - Cytotoxic T-lymphocyte-associated protein

CTNNB - Catenin beta

CXCR - C-X-C motif chemokine receptor

D - Diversity

DENND - Differentially expresed in normal and neoplastic cells domain-containing protein

DHDDS - Dehydrodolichyl diphosphate synthase complex subunit

DLA - Dog leukocyte antigen

DN - Double negative

DNA - Deoxyribonucleic acid

DNMT - DNA methyltransferase

DNTT - DNA Nucleotidylexotransferase

DP - Double positive

DUS4L - Dihydrouridine synthase 4 like

DUSP - Dual specificity phosphatase

EATL - Enteropathy associated T-cell lymphoma

EEF1A1 - Eukaryotic translation elongation factor 1 alpha 1

EGFR - Epidermal growth factor receptor

EIF - Eukaryotic translation initiation factor

ELK - ETS transcription factor

EML - Echinoderm microtubule associated protein like

ENKTCL - Extranodal NK-/T-cell lymphoma

EOMES - Eomesodermin

EP300 - E1A-associated protein p300

ERBB - Erythroblastic leukemia viral oncogene homolog

ERK - Extracellular signal-regulated kinase

ETV - ETS variant transcription factor

FAM - Fluorescein amidites

FANCA - FA complementation group A

FASL - Fas ligand

FAU - FAU ubiquitin like and ribosomal protein S30 fusion

FC - Fold change

FDR - False discovery rate

FISH - Fluorescence in situ hybridization

FLT - FMS-like tyrosine kinase

FOXO - Forkhead box

FOXP - Forkhead box protein

FSTL - Follistatin like

FYN - FYN Proto-oncogene

G - Guanine

GATA - GATA binding protein

GATD3A - Glutamine amidotransferase class 1 domain containing 3

GATK - Genome Analysis Toolkit

GEM - Gel bead-in-emulsion

GEO - Gene Expression Omnibus

GFF - General feature format

GIMAP - GTPase of the Immunity-Associated Protein Family Member

GLUT - Glucose transporter type

GNAS - Guanine Nucleotide-binding protein, alpha stimulating

GRB - Growth factor receptor-bound protein

GSEA - Gene set enrichment analysis

GSVA - Gene set variation analysis

GTF - General transfer format

GZMA - Granzyme A

HAVCR - Hepatitis A virus cellular receptor 1

HiFi - High fidelity

HLA - Human leukocyte antigen

HNGC - HUGO Gene Nomenclature

HOPX - Homeodomain-only protein

HSPD - Heat shock protein family D

HSTL - Hepatosplenic T-cell lymphoma

IACUC - Institutional Animal Care and Use Committee

IBTK - Inhibitor of Bruton tyrosine kinase

ICOS - Inducible T cell costimulator

IFN - Interferon

Ig - Immunoglobulin

IHC - Immunohistochemistry

IKZF - IKAROS family zinc finger

IL - Interleukin

IMHA - Immune mediated hemolytic anemia

IMTP - Immune mediated thrombocytopenia

INDEL - Insertion-deletion

INPP4B - Inositol polyphosphate 4-phosphatase type II

IP - Interacting protein

IRF - Interferon regulatory factor

IRS - Insulin receptor substrate

ISP - Immature single positive

ITGA - Integrin subunit alpha

ITK - Interleukin-2-inducible T-cell kinase

ITM - Integral membrane protein

J - Junctional

JAK - Janus kinase

JAZF - Juxtaposed with another zinc finger

JPT - Jupiter microtubule associated homolog

JUN - Jun Proto-oncogene

kb - Kilobase

KCND2 - Potassium voltage-gated channel subfamily D

KDR - Kinase insert domain receptor

KHDRBS - KH RNA binding domain containing, signal transduction associated

KIT - KIT Proto-oncogene

KLF - Krüppel-like factor

KLRB - Killer cell lectin-like receptor subfamily B

KLRK1 - Killer cell lectin like receptor 1

KMT - Lysine-specific methyltransferase

KRAS - Kirsten rat sarcoma virus

LATS - Large tumor suppressor kinase

LB - Lysogeny Broth

LCK - LCK Proto-oncogene

LMO - LIM domain only

lncRNA - Long non-coding RNA

LOX - Lysyl oxidase

LTA4H - Leukotriene A4 hydrolase

LTB - Lymphotoxin beta

LYZ - Lysozyme

MAF - Mutation annotation format

MAML - Mastermind like transcriptional coactivator

MAP2K - Mitogen-activated protein kinase kinase

MAPK - Mitogen-activated protein kinase

MCM - Minichromosome maintenance complex component

MEITL - Monomorphic epitheliotropic intestinal T-cell lymphoma

MET - Mesenchymal epithelial transition

MHC - Major histocompatibility complex

MRC - Mannose receptor, C-type

mRNA - Messenger RNA

MROH1 - Maestro heat like repeat family member

MSH - MutS homolog

MSigDB - Molecular Signatures Database

MTOR - Mammalian target of rapamycin

MYC - MYC Proto-oncogene

MYH - Myosin heavy chain

NA - Not applicable

NCBI - National Center for Biotechnology Information

NCL - Nucleolin

NCOR - Nuclear receptor co-repressor

NEDD4L - Neural precursor cell expressed developmentally downregulated gene 4-like

NES - Normalized enrichment score

NF-kB - Nuclear factor kappa-light-chain-enhancer of activated B cells

NK - Natural killer

NLRP - NOD-like receptor family pyrin domain containing

NOS - Not otherwise specified

NOTCH - Neurogenic locus notch homolog protein

NPR - Natriuretic peptide receptor

NT5DC - 5'-nucleotidase domain containing

nTFHL - Nodal T-follicular helper cell lymphoma

nTFHL-F - nTFHL-follicular type

ORF - Open reading frame

padj - Benjamini-Hochberg adjusted p-value

PAK - p21-activated kinase

PARP - Poly(ADP-Ribose) polymerase family member

PARR - PCR for antigen receptor rearrangement

PAX - Paired box

PAX - Paired box

PCA - Principal component analysis

PC-ALCL - Primary cutaneous ALCL

PCGF - Polycomb group ring finger

PCR - Polymerase chain reaction

PD - Programmed cell death

PDE - Phosphodiesterase

PDGFR - Platelet-derived growth factor receptor

PER - Period circadian regulator

PI3K - Phosphatidylinositol-3-kinase

PLCG - Phospholipase C, gamma

PLEKH - Pleckstrin homology domain containing

PSMA - Proteasome subunit alpha type

PTCL - Peripheral T-cell lymphoma

PTEN - Phosphatase and tensin homolog

PTGS - Prostaglandin-endoperoxide synthase

PTMA - Prothymosin alpha

PTPRF - Protein tyrosine phosphatase receptor type F

QV - Quality value

Rac - Ras-related C3 botulinum toxin substrate

RAG - Recombination-activating gene

RAS - Rat sarcoma

RBPMS - RNA binding protein, mRNA processing factor

REV3L - Protein reversionless 3-like

RGS - Regulator of G protein signaling

RIPOR - RHO family interacting cell polarization regulator

RNA - Ribonucleic acid

RNA-seq - RNA sequencing

ROR - Retinoic acid-related orphan receptor

RP - Ribosomal protein

RT - Reverse transcription

RUNX - Runt-related transcription factor

SATB - Special AT-rich sequence binding-protein

SCA - Stem cell antigen

SCML - Scm polycomb group protein like

SELL - L-selectin

SERPINB - Serine protease inhibitors family B

SH2D - SH2 domain–containing protein

SLC2A4 - Solute carrier family 2, facilitated glucose transporter member 4

SLC45A3 - Solute carrier family 45 member 3

SMAD - Suppressor of mothers against decapentaplegic

SMO - Smoothened, Frizzled class receptor

SMRT - Single-molecule real-time

SNP - Single nucleotide polymorphism

SOX - SRY-box transcription factor

SP - Single positive

SPOCK2 - Secreted protein acidic and rich in cysteine (SPARC) / osteonectin

SRSF - Serine and arginine-rich splicing factor

STAG2 - STAG2 cohesin complex component

STAT - Signal transducer and activator of transcription

STIL - SCL/TAL1-interrupting locus

STK - Serine/threonine Kinase

SUZ12 - SUZ12 polycomb repressive complex 2 subunit

SYK - Spleen associated tyrosine kinase

T - Thymine

TAL1 - T-Cell Acute Lymphocytic Leukemia

T-ALL - T-cell acute lymphoblastic leukemia

TBC1D - TBC1 Domain Family Member

TBX - T-box transcription factor

TCF - T cell factor

TCR - T-cell receptor

TE - Tris-EDTA

TET - Tet methylcytosine dioxygenase

TFAM - Transcription factor A, mitochondrial

Tfh - T follicular helper

TGF - Transforming growth factor

Th - T helper

THEMIS - Thymocyte selection associated

THPOK - T-helper-inducing POZ-Kruppel factor

TIM - T-cell immunoglobulin and mucine-domain containing

TMPRSS - Transmembrane serine protease

TNF - Tumor necrosis factor

TNFAIP - Tumor necrosis factor alpha induced protein

TNFRSF - Tumor necrosis factor receptor superfamily member

TOX - Thymocyte selection associated high mobility group box

TP - Tumor protein

TPD52L - Tumor protein D52 like

TPRG - Tumor protein p63 regulated

TRAF - Tumor necrosis factor receptor-associated factor

Treg - Regulatory T cells

TRIB - Tribbles pseudokinase

TRIM - Tripartite motif-containing

TROP - Tumor-associated calcium signal transducer

TZL - T-zone lymphoma

U - Uracil

UMAP - Uniform manifold approximation and projection

UMI - Unique molecular identifier

UTR - Untranslated region

V - Variable

VAV - Vav guanine nucleotide exchange factor

VBP - Von Hippel-Lindau-binding protein

VCF - Variant call files

VLA - Very late antigen

WES - Whole exome sequencing

WHO - World Health Organization

XPO - Exportin

YWHAZ - Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Zeta

ZAP - Zeta chain of T-cell receptor associated protein kinase

ZMW - Zero-mode waveguides